UNIVERSITY
OF FLORIDA
LIBRARIES

Digitized by the Internet Archive
in 2013

http://archive.org/details/workeffort00ryan

WORK AND EFFORT

THE PSYCHOLOGY OF PRODUCTION

By

THOMAS ARTHUR RYAN

ASSOCIATE PROFESSOR OF PSYCHOLOGY

CORNELL UNIVERSITY

THE RONALD PRESS COMPANY · NEW YORK

658.5
R989w

ENGINEERING
SCIENCES
LIBRARY

Copyright, 1947, by
THE RONALD PRESS COMPANY

All Rights Reserved

The text of this publication or any part thereof may not be reproduced in any manner whatsoever without permission in writing from the publisher.

PRINTED IN THE UNITED STATES OF AMERICA

To

MADISON BENTLEY

PREFACE

This book provides a systematic survey of psychological investigations concerned with the productivity of men and women at work. It attempts to present an evaluated account of the knowledge that has been accumulating for many years, to give a survey of current progress in this field of investigation, and to put the reader in position to realize the need for future development in researches of this character.

Among the problems treated are training and learning, the control of accidents, establishment of pay levels for various jobs, and the design of efficient methods of work. Motivation of the worker is considered in its relation to problems of efficiency. Because all these topics have a common core, they are treated broadly and with regard to their general rather than specific implications.

The book presents conclusions drawn from research in various fields. These include a study of the factors of light, heat, and ventilation, the control of monotony and fatigue, and the provision of conditions that make for maximum output without excessive effort or fatigue. All this information is based upon reliable experimental and statistical techniques, to the end that facts, and interpretations which are close to the facts, may be provided for the use of industrial managers and supervisors, personnel directors, and psychologists—in short, for all those who are concerned with the setting of standards, with installing merit rating and job evaluation systems, and with improving the morale and efficiency of workers by the use of tested scientific methods.

In addition to presenting material basically related to the topics mentioned above, the book also aims to acquaint the reader with the methods of research that led to the findings set forth. This has been done so that the reader may determine for himself just how the psychologist arrives at any given conclusion. Such a manner of presentation should also appeal to students of applied psychology who plan to go into engineering, managerial, personnel, or other related lines of work.

In acknowledging obligations to those who aided the author in writing this book, the first is inadequately recognized by dedicating this work to Professor Madison Bentley. The author is indebted to

him for not only tuition but also sage advice and helpful guidance during the author's initiation into the field of psychology.

Another important contribution to the work was made by the College of Engineering at Cornell University. Some years ago the author was asked to join the faculty of that College in order to discover, if possible, what contributions a psychologist might be able to make to the solution of problems that arise in engineering work. Dean S. C. Hollister, responsible for giving the author time and opportunity to become acquainted with the engineering problems involved, displayed an active interest in furthering studies in this field. Other members of the College faculty contributed in many ways, especially Professors John R. Bangs, Jr., C. L. Cottrell, D. Dropkin, H. J. Loberg, C. I. Millard, and E. M. Strong.

Colleagues in the Department of Psychology also gave much help. Professor M. E. Bitterman read and helpfully criticized the whole manuscript—some parts of it several times at various stages of writing. Portions of the material were also read by Professors H. P. Weld, H. S. Liddell, and F. L. Marcuse. A former colleague, Professor John G. Jenkins, now of the University of Maryland, first introduced the writer to the field covered by this book. Those who know Professor Jenkins' work will doubtless find many influences, particularly on the general point of view expressed.

The author is grateful to the many writers and publishers who kindly allowed quotations, charts, and tables to be taken from their books and articles. These permissions are acknowledged wherever the material appears in the text.

Last of the author's obligations—but far from the least of them—he owes to his wife, Dr. Mary Shaw Ryan, who supplied criticism of psychological content and literary expression, as well as constant encouragement throughout all the time this book was being written.

T. A. R.

Ithaca, New York
September 2, 1947

CONTENTS

CHAPTER		PAGE
1	Introduction	3
2	The Meaning of Efficiency or Economy of Work	20
3	Measuring the Cost of Work: I. Metabolism and Muscular Fatigue	33
4	Measuring the Cost of Work: II. Fatigue in Sedentary Work	63
5	Measuring the Cost of Work: III. Physiological Tests of Effort and Fatigue in Sedentary Work; Statistical Indices of Efficiency	95
6	Specific Factors Governing Efficiency and Productivity	124
7	Work Methods and Efficiency	157
8	Incentives and Motives	169
9	"Nervous" Fatigue and Boredom	188
10	Time Standards and Rate Setting	209
11	Merit Rating and Job Evaluation	237
12	Psychological Problems in Accident Control	253
13	Skill and Practice	297
	Index of Names	313
	Index of Subjects	317

ILLUSTRATIONS

FIGURE		PAGE
1	Scatter Diagrams Representing Varying Degrees of Correlation Between Grades in Physics and in Calculus	14
2	Relation Between Output and Feelings in Four Forms of Mental Work	46
3	Finger Ergograph as Designed by Mosso	49
4	Records of Two Ergograms	50
5	Skin Resistance and Muscle Potentials Following Ergographic Work	59
6	Hypothetical Effects of Fatigue and Practice	67
7	Composite Hourly Performance Curves for Two Metalworking Plants	71
8	Composite Hourly Performance for Different Classes of Work (Eight-hour Metalworking Plant)	72
9	Hourly Performance Curves for Visual Inspection of Cartridge Shells	74
10	Test Performance Following Three Nights of Diminished Sleep (Dotting Test)	82
11	Test Performance Following Three Nights of Diminished Sleep (Associated Words)	83
12	Action Potentials During Rest and Work (Reading), With and Without Distraction	105
13	The Relation Between Critical Flicker-Fusion Frequency and Brightness	115
14	Key-Pressure in Work During Quiet and Noise	126
15	Average Magnitude of Action Potentials on Successive Days of Stimulation	127
16	American Society of Heating and Ventilating Engineers Comfort Chart for Still Air	136
17	Practice Curves in Computation for Different Instructions	172
18	Performance in Computation Under Two Instructions	173
19	Average Hourly Output in Western Electric Company Test Room Studies	180

FIGURE PAGE

20 Effects of Incentives Upon Different Kinds of Work . . . 186
21 Hypothetical Work Curves 198
22 Production Curves for a Repetitious Task—Individual Curves Compared with Worker's Reports on Boredom 200
23 Production Curves for One Week of Repetitious Work—One Worker Who Considers the Work Very Boring 201
24 Fatigue and Personal Allowance Table 226
25 Standard Time Values for "Place" Corrected for Transport Distances 230–231
26 Example of a Typical Merit Rating Form 239
27a Suggested Form of Employee Appraisal 247
27b Inventory of Personnel 248
28 Hourly Incidence of Accidents in Various Industries . . . 258
29 Accidents and Production in Eight-hour and Ten-hour Metalworking Plants 261
30 Accident Frequency in Relation to Temperature (Shell Factory) . 263
31 Accident Disability in Relation to Temperature (Coal Mining) . 264
32 Driver Testing Apparatus with Moving Road Scene . . . 271
33 Accident Rates of Dockyard Apprentices, Related to Test Scores ("Aestheto-Kinetic Tests") 274
34 Relation of Accidents to Scores on Driver Tests 277
35 Accident Reduction Following the Use of Special Tests (Accident Record of 144 Accident Repeaters) 284

TABLES

TABLE		PAGE
1	Range of Correlations Obtained by Chance for Varying Size of Group, Assuming That No Correlation Exists in the Population from Which the Group Is Drawn	16
2	Types of Muscular Work	35
3	Rough Computation of Mechanical Efficiency	37
4	Cost of Work in Lifting Loads by Crank	38
5	Relative Energy Cost of Pulling vs. Pushing	39
6	Energy Expended in Human Locomotion	40
7	Kinds of Sedentary (Mental) Work	64
8	Change in Performance from Morning to Evening	84
9	Proportion of Days Showing Decreased Performance	85
10	Zones of Maximum Comfort	137
11	Effects of Increased Illumination Upon Industrial Output	139
12	Critical Levels of Illumination Beyond Which There Is Little Change in Output	139
13	Influence of Level of Illumination	140
14	Load on Fingers During Typing	166
15	Rank Assigned Various Factors in Morale by Employers and Employees	177
16	Reasons Assigned by 157 Young Men, High School Graduates, for Preferring One Job Rather Than Another	177
17	Average Boredom Rating of 68 Workers According to Types of Work Curves	199
18	Average Scores in a Test of Intelligence of Individuals Who Received the Highest and Lowest Assessments for Boredom	204
19	Incidence of Boredom Symptoms in the Case of the Same Group of Workers Employed for Monthly Periods in Different Processes	207
20	Leveling Factors	215
21	To Illustrate the Application of Leveling Factors	216
22	Time Study of Operators' Performance in a Folding Job	220
23	Mean Cycle-Time, Standard Deviation, and Range	234

TABLE		PAGE
24	Hypothetical Ratings of Three Workers with Equivalent Total Scores	242
25	Factors in Job Evaluation	251
26	Total Accidents: December 7, 1941 to August 14, 1945	253
27	Comparison of Morning and Afternoon Accidents	259
28	Effect of Length of Working Day Upon Increase of Accidents in the Afternoon	259
29	Variation in the Average Number of Hospital Visits per Year Among Employees on Eleven Different Jobs	265
30	Distribution of Automobile Accidents in a Random Sample	268
31	Hypothetical Analysis of Accident Distribution with Two Liability Groups	269
32	Fletcher's Comparison of Test Performance of "Good" and "Poor" Drivers	273
33	Mean Percentage Accident Rate Over the Whole Period of Exposure of Groups Selected by Different Methods	275
34	Effect on the Subsequent Accident Rate of Removing Drivers with Many Accidents in Their First Year or Removing Drivers with Poor Aesthetokinetic Co-ordination	275
35	Vision and Intersection-Type Accidents	276
36	Correlations with Accident Rate	279
37	Re-educational Effect of Driver Test Clinics	282
38	Product-Moment Correlation Coefficients Between Different Types of Accident	287
39	Personal Accidents of Drivers for a Public Utility Company Compared with Motor Vehicle Accidents of the Same Drivers	289
40	Kinds of Learning	302

WORK AND EFFORT

Chapter 1

INTRODUCTION

Supervisors and managers of industry, teachers, students, and workers, all have an interest in the productivity and efficiency of man at work. They are interested in providing conditions which are conducive to maximum output without excessive effort or fatigue. They wish to know how important such factors as light, heat, ventilation, hours of work, may be in determining performance. They would like to know how to control monotony and fatigue, and how to design methods of work which are easier and more rapid than those which have been used in the past. They inquire most frequently of all about such topics as morale and motivation.

This book is a survey of psychological investigations of these problems. It aims to give the student of applied psychology an understanding of the basic knowledge in this field which has been accumulating during the past few decades. In the relatively short time since research upon these problems began, much has been discovered, but naturally much remains to be done. We hope that the student will not only gain a knowledge of our current progress, but that he will also come to a realization of the future developments which are necessary in this field before some of the important problems of practice can be dealt with on a sound basis.

We consider it to be the task of the applied psychologist to provide information which is based upon reliable experimental and statistical techniques—in short, to provide *facts* and interpretations which are close to the facts. When the psychologist steps beyond these limits, he should make it very clear that he has done so—that he is now guessing, using "common sense" or hunches.

Complete Solutions vs. Solutions Limited by the Available Facts.—The problems with which this book will deal affect every human being. Because of their importance it is natural for the writer on these subjects to be discontented if he is unable to give a well-rounded, easy, and "helpful" answer to many of the questions which arise in the factory, office, school, or home. He is subjected to

great pressure from personnel men, supervisors, and teachers to "give an answer we can use right away." The temptation to go far beyond the established facts in meeting such a request is very great, especially as the layman is not likely to look closely at the evidence upon which the recommendations are based, but one we should and will try to avoid.

The progress of applied psychology would be much more rapid if we disappointed industrial managers and supervisors more frequently, pointing out that we would have to guess, with little guarantee that our guesses would be better than anyone else's. Industry has had to put engineers to work solving its physical problems and filling the gaps in available information. There is no reason why industrial and business men should expect the universities to fill in these gaps in psychological knowledge without support, and they would not do so if they realized what is involved in the questions they raise.

Let us take a simple problem from our own field to illustrate the point. Suppose we are asked to recommend the best temperature to be maintained in an office building, from the point of view of efficient mental work. Someone has mysteriously arrived at the conclusion that 68 °, 50 per cent relative humidity, is the best and most healthful condition for sedentary people. That recommendation is given again and again without any reference to the fact that there is no discoverable and adequate experimental basis for the recommendation. Would it not be better to point this out, and refuse to give a recommendation until those who ask the question are willing to support an experimental program for seeking an answer to it?

The point of view expressed here is not always accepted, even by psychologists. There are those who feel that the psychologist should try to give some definite recommendation even in those problems in which it is necessary to "go out on a limb" and make statements based upon personal opinion. The justification which they use is that the business man, the teacher, or the industrialist must make a decision about the question. If the psychologist will not give him a definite answer, he will use his own judgment, or he may become the prey of quacks, or other unscrupulous persons who claim to solve all problems by some method completely unfounded in scientific investigation. Some psychologists, concerned about this state of affairs, say, in effect, "Would it not be better for the psychologist to try to give an answer to the best of his ability? The psychologist may make mistakes, but his conclusions will be based upon the best available evidence, and certainly would not be based upon theories which are entirely false and misleading."

This argument may have force when it refers to the functions of the psychologist in private consultation on some specific problem, although even there it is questionable whether the opinions of the psychologist are to be given more weight than those of an experienced supervisor or executive. In such consultations, the main advantage which the psychologist has is that he is outside the situation and can look at the problem more impersonally. In many problems, the actual facts which he can use in drawing his conclusions will be less important than this impersonal, objective approach.

A book aiming to present some general field of applied psychology, and attempting to follow the procedure we have just outlined (the procedure of being "practical" at all costs), would have several serious defects. If the book is limited to recommendations for practice, the reader has no way of judging which recommendations are based upon a wealth of research information and which are based almost entirely upon guessing. Just as important as this defect is the fact that the book must be either an encyclopedia of hundreds of specific problems and their solution, or it must attempt to present so-called general "principles" which are presumed to be general enough to apply in a variety of situations. In the former case, the possibilities of specific situations and problems are never exhausted, and in the latter case the general "principles" themselves are products of judgment, and further judgment must be exercised in their application.

It should be just as easy for the student of psychology or the practical man of affairs to learn something of the facts and methods by which the psychologist works as it is to learn the recommendations found in books of either of the types described in the preceding paragraph. The results should be more effective in practice. The man of affairs would be less tempted to succumb to the salesmanship of the quacks if he knew something of the errors in their methods.

Need for Understanding the Methodology of Psychological Research.—Even if we limit ourselves to a discussion of facts and methods whose value has been the subject of controlled research, there is still another choice which must be made. We may present only the conclusions to be drawn from the research which has been done in each field, or we may try to show the reader how those conclusions were derived by giving him a summary of the research methods and results upon which the conclusions are based. The first alternative is frequently adopted in texts upon applied psychology or personnel management, apparently on the assumption that those who are not planning to carry on research of their own do not need the background of methods.

There are several reasons why we disagree with this latter point of view, and why we believe that some treatment of the methodology of research is essential, even if the presentation is not specifically designed for the student who will engage in research of his own. Even the industrial manager or engineer who is immediately concerned with practice would profit, we believe, by a knowledge of research methods and specific results in applied psychology. Perhaps he would profit even more from this kind of information than he would from general conclusions and recommendations. Here, in brief, are some of the reasons:

1. Conclusions are often drawn with reference to a certain state of affairs, or conditions which exist now. When new situations arise, the conclusions may not be applicable at all, while certain of the facts would be reinterpreted with reference to the new situation.
2. The practical man of affairs may feel that the psychologist is being too cautious and limiting his conclusions too much—hedging them about with qualifications and provisos. If the facts themselves are available, he should see that the only way in which the conclusions can be broadened with assurance is by more extensive research. If the reader wishes to draw wider conclusions he may do so—at his own risk and with better appreciation of the risk.
3. If the layman can arrive at an appreciation of the care which is necessary in research to arrive at reliable and valid results, he will become aware of the unreliability of common sense and experience as means of solving practical psychological problems. For example, if he sees how important such factors as suggestion may be in determining the effects of any given factor, he will look more critically at his own experience and at the anecdotes which he hears from others.
4. Many students of applied psychology, while they plan to go into engineering, managerial, personnel, or other similar types of work, will frequently become involved in some way in psychological research. They may have to deal with some problems themselves, they may have to decide when the services of a specialist in a particular kind of research are needed, or they may have to evaluate the results of research carried out by others.

In this book, therefore, we shall take the position that one cannot get the most out of the results of applied psychological researches without knowing something of the method by which the facts were

derived. It is not enough that the writer aims to draw his conclusions from facts, rather than from opinion and speculation. The scope and meaning of the generalities are not clear without the research background. The engineer who has to work in a technological field is expected to know something of the fundamental research methods underlying the procedures which he applies. Similarly, one who wishes to deal with practical problems of human behavior in a technological fashion should have a background in psychological research.

Major Problems of Work

Before going into detailed analysis of any one phase of work and efficiency, it would be well to investigate the scope of the field which we propose to study. Let us consider the broad problems involved, and their general meaning. It will be seen that these problems, which appear at first quite distinct, have a common core which will be made the central theme of this book.

1. Problems of Efficiency.—Here are included such questions as these: "What is the best way to learn to operate a typewriter?" "Is it better to use a great deal of specialization, or let each worker perform a number of tasks?" "How frequently should the workers on a certain job rest, and what should they do during the period of rest?" "Would it be worth while installing better lights, an air-conditioning system?" "Would I be just as well off with six hours of sleep every night instead of eight or nine—am I wasting time in sleeping?"

This list could be extended indefinitely, of course. It is also evident that many of these questions are complex, and so cannot be answered fully as yet. Before we can even seek an answer, however, it will be necessary to clarify the meaning of many such questions. When we ask, "What is the best way to do a certain thing?" we may get in reply a counter-question, "The best way for *whom*?" Would the answer be the same from the point of view of the worker, of the employer, or of the community at large? Even more generally, we may ask, "The best way from what point of view?" Do we want to seek the method which is the best from the point of view of rate of production, amount of fatigue per unit of work, or interest in the work? In brief, when we consider questions of efficiency, we shall have to consider what we mean by efficiency, and also whether different points of view will lead to the same result. Until we do so, we shall not know what we are looking for.

We shall consider efficiency in terms of a relationship between the cost of work to the individual (effort, fatigue, energy expended, etc.)

and the outcome. If cost of work is constant while the output increases, we shall speak of an increase in efficiency. If the output is increased but the cost is unknown, we shall not be justified in speaking of an increase in efficiency.

Terms like *effort* and *fatigue* are very familiar and are readily understood in the above definition. We shall find, however, that the measurement of effort and fatigue requires something more than a general understanding of the terms. In fact, these are key problems of the psychology of work.

2. Problems of Motivation.—One group of questions under this heading concerns individual differences in motivation: "Why does A get so bored with his work, while B is much more interested?" "What can be done to increase A's interest, or how can another job be found which will appeal to him more?" Another set of questions relates to *group* or *average* motivation: "What inducements can be offered to make a suggestion-system more effective?" "What kind of campaign will induce workers to be more careful about leaving things about the floor?" "How can the productivity of this plant be increased without a greater working force—*i.e.,* how can the workers be induced to work harder? What kind of bonus system would be most effective?" "How can we persuade the stenographers to remain on their jobs and stop looking for factory jobs?"

There is a very simple relationship between problems of efficiency and problems of motivation—at least in the conception, if not in the solution, of the problems. In problems of motivation, we are basically concerned with raising the level of effort that the worker puts into his work, and thus increasing his productivity. In the problems of efficiency which were discussed above, the problem is to increase the results which are obtained from a given amount of effort—to see that the effort is expended more economically.

3. Special Problems.—There are a number of problems which are already included under the headings of efficiency and motivation. Because of their unique characteristics or their special importance, however, they are usually given separate mention:

(*a*) Training and learning.
(*b*) Accident control.
(*c*) Establishing relative pay levels for various jobs.
(*d*) Selecting workers.

Of these special topics, the last, selecting workers, has received so much elaboration in recent years, and there are so many specialized

techniques available, that in this book we shall mention it only in passing. Our interest will be mainly that of relating the problem to other aspects of efficiency and motivation.

Topic (*c*) is another instance of the great importance of measurement of effort. Here we come immediately to the question, "For what is the worker being paid?" It is obvious that he is being paid for his skill, knowledge, and the effort he expends upon his work. In a sense, he is paid for past and present effort, plus the aptitude which he brings with him from his family inheritance and early upbringing. This is an extremely simplified statement, but when we come to analyze the problem of establishing pay levels we shall see that the measurement of effort is involved in a number of complex ways. In fact, there are several methods that are spoken of as "scientific" in this field which do not deserve that label because they are not based upon any sound method of evaluating effort.

Methods of Gathering Facts and Determining Their Interrelationships

As soon as we enter any field of psychology we are confronted with a maze of opinions, statements, and claims. What criteria shall we use in sorting these statements and in determining which are worthy of serious consideration, which are untrue, and which require further checking? For example, which of the following statements would you accept as factual?

1. Women are better adapted to repetitious work than men.
2. Women cannot work as long hours as men.
3. Persons of normal or high intelligence are very unlikely to do well at simple manual and mechanical tasks.
4. The primary incentive factor in industry today is the pay rate.
5. Loss of a night's sleep has a serious effect upon performance of a job the next day.

Some of these statements refer to a state of affairs which is supposed to exist. In other cases (for example, Statement No. 5) a cause-and-effect relationship is involved. In the latter case we must determine a relationship between the facts, as well as establish the facts themselves. Let us first consider what would be necessary to establish our facts, and then come back to the further problems of cause-and-effect or other relationships between the facts.

The statement that women are better adapted than men to repetitious work is based upon the opinions and reports of personnel

administrators and supervisors. Does this justify its acceptance as an established fact? It is no reflection upon the veracity of these observers if we suggest that they might be wrong in their opinion. It it not a question of honesty, but rather a question as to whether it is possible to arrive at such a conclusion on the basis of common, everyday experience. Such statements have been made with assurance in the past and have later been proved to be wrong. We no longer believe that the earth is flat, but present-day psychological beliefs are in many ways comparable to the knowledge of geography and astronomy which was available to Columbus.

A very recent example illustrates the danger of trusting to this method of common observation for our "facts." Management of a certain concern stated that in that organization the worst day of the week for absenteeism was the day after payday. Study of the statistics of absences of that organization, however, showed that Monday had a considerably higher rate of absences than the day that had been called the worst of the week. If a mistake of this kind, dealing with a simple question of frequency of absence can be made, how many mistakes must appear when we deal with something less tangible; such, for example, as the degree of adjustment to repetitious work!

This example of absentee rate brings out clearly what is needed to establish the correctness of other statements—measurement or counting instead of unreliable estimation. We must have some measure of degree of adjustment to routine work, or we must at least be able to classify men and women into groups (high, medium, or low). This measure of adjustment must be applied to a *representative sample* of both men and women, working on a variety of jobs. In order to make the general statement, we should have to be certain that we have studied representative routine jobs and, of course, the same jobs for both men and women as well as a representative group of men and women. Finally, we need to know the *statistical significance* of the difference, as well as the size of the difference and its practical importance. We need, in short, a statistical analysis of the results.

Statistical and Practical Significance.—To a scientific worker, statistical analysis means much more than mere computation of averages. Much more important is the interpretation of the results from the point of view of *chance*. When we say that a result is *statistically significant,* we mean that we have good reason to accept the result as more than mere luck, that we could repeat the experiment and get similar results again, that the result is *dependable*.

Suppose, for example, that we had studied 100 men and 100 women working upon a dozen different routine kinds of work. We might get any one of the following results:

(*a*) A large but *in*significant difference in adaptation to work. That is, while the difference between the averages appears substantial, there is so much individual variation among both men and women that we should have to have a much larger group in order to determine whether any real difference exists.

(*b*) A large, significant difference. That is a difference which could be depended upon, which would again appear if the study were to be repeated.

(*c*) A small, significant difference. This would also indicate that there is some real difference between men and women in this respect, but that there would be a considerable amount of overlapping. A substantial number of men would be better than many of the women, and, conversely, many women would fall below the men's average.

A significant difference is not necessarily of great practical usefulness. In case (*c*) above, we might have only a slightly better chance of getting a satisfactory worker if we hired a woman instead of a man. Statistical significance depends not only upon the size of difference but also upon the number of cases studied and the amount of individual variation within the groups. A very slight difference between the averages of two groups might be significant if there were a million people in each group, although individual members of each group are spread over a wide range.

We must establish the statistical significance of any result before we are justified in drawing any conclusions at all. Once it is established, we must still determine the practical meaning of the result.

Two common errors which are found in reports of investigations of practical psychological problems involve the following points: First, there are many investigations carried out by practical men of affairs who believe that statistical analysis ends with the computation of averages. Obviously, in the light of what has just been said, no conclusions at all are possible. The second error is sometimes overlooked even by writers who have some knowledge of statistical method. This consists in determining the statistical significance, but failing to evaluate the findings from the point of view of practical applicability. The statement quoted in the list above, to the effect that intelligence is related to ability to adjust to repetitious work, is based upon such a failure. *Barely* significant and quite small differ-

ences were obtained in the investigation. Yet some books on personnel management regard these results as showing the practical value of selecting routine workers on the basis of intelligence level. Actually, the overlapping between the groups was so great, and the number of different jobs studied was so small, that the significance of the difference is not enough to establish intelligence tests as a useful selective tool.

A simple numerical illustration should make the latter point clear. In a hypothetical study, the proportions of men and women who are seriously affected by boredom in their work are as follows:

	Per Cent Seriously Bored
Men	30
Women	25

If the two groups are sufficiently large, this difference could be highly significant in a statistical sense. This means that further studies with large numbers of individuals would also obtain differences in the same direction. The decision to eliminate men from this kind of work, however, would result in only a small decrease in the percentage of workers who are bored with the work. Many other considerations would have to be taken into account before we could decide whether the elimination of male workers would have any real advantage at all.

Relationships Between Variables.—Two general kinds of interrelationship between events are of importance in applied psychology. We may be interested in a relationship between two variables because we wish to predict one of them from the other. Thus, we may wish to predict performance in a certain college course from an individual's performance in a previous course. In this case we are not concerned with the reasons for the relationship. We are satisfied if some predictable relationship exists. In other cases, however, we are interested in the relationship because we want to know what kind of relationship exists. We wish to know *in what way* one variable depends upon the other. Thus, if performance in a course in physics is predictable from performance in an earlier course in calculus, we may want to know whether the dependence is due to the need for the material learned in the calculus course, or whether the calculus grade is merely an indication of a special aptitude of the student which determines good physics grades regardless of whether he has studied calculus or not. In other words, does performance in physics *depend upon* the calculus learned, or are the two variables related through

some third factor? In both cases we could predict the second variable from the first, but our interpretation of the facts and other applications would differ in the two cases.

We have a variety of measures of statistical relationships between variables at our disposal. None of these, however, can do more than measure the predictability of the relationship; they do not indicate the nature of the dependence, the presence or absence of third factors, or anything of that sort. They merely tell us how accurately one variable can be predicted from the other.

Determining the predictive relationship is, nevertheless, a step which is useful in itself, and which must be part of any analysis of the dependence of variables. It will be well, therefore, to become familiar with, or to review, as the case may be, the important tools that are used for this purpose.

The most commonly used measure of predictability is the *coefficient of correlation.* Its use is so widespread that no one can hope to read very far in serious psychological literature without at least a general understanding of its meaning. The coefficient of correlation (commonly designated by the letter r) is an index which varies from -1.00 to $+1.00$. To return to the illustration of the relationship between calculus and physics, we should investigate that relationship by obtaining the records in both courses of a large number of students. We might plot the points representing the individual students upon a graph like those shown in Figure 1, where the x axis represents performance in calculus, and the y axis represents performance in physics. If all the points fell on a straight line like that in Figure 1*a,* we would have a perfect correlation (1.00), and we would be able to predict physics grades exactly, once the calculus grade were known. If the points scattered, as in Figure 1*b,* we could not predict from one variable to the other any better than we could if we drew numbers out of a hat. This is the condition expressed as a coefficient of 0.00. A minus sign for a coefficient merely indicates an inverse relationship. In this case a negative correlation would mean that physics grades are lower, the higher the student's grade in calculus. Figure 1*c* represents the "scatter diagram" which would be obtained in a case of -1.00 correlation.

Correlations of 1.00 are never obtained in practice because of variations due to chance (any unpredictable variable) and those due to other factors which are not under control or study. For example, the number of hours of work the student is taking at the time of the course in physics may not be the same as that taken at the time of the course in calculus. We must, therefore, be able to deal with cases in

which there is a general, but not exact, relationship between the variables. This situation is expressed by a correlation coefficient between 0 and 1 (with either a positive or negative sign, depending upon the direction of the relationship). Figure 1*d* represents results which would be expressed by such a coefficient.

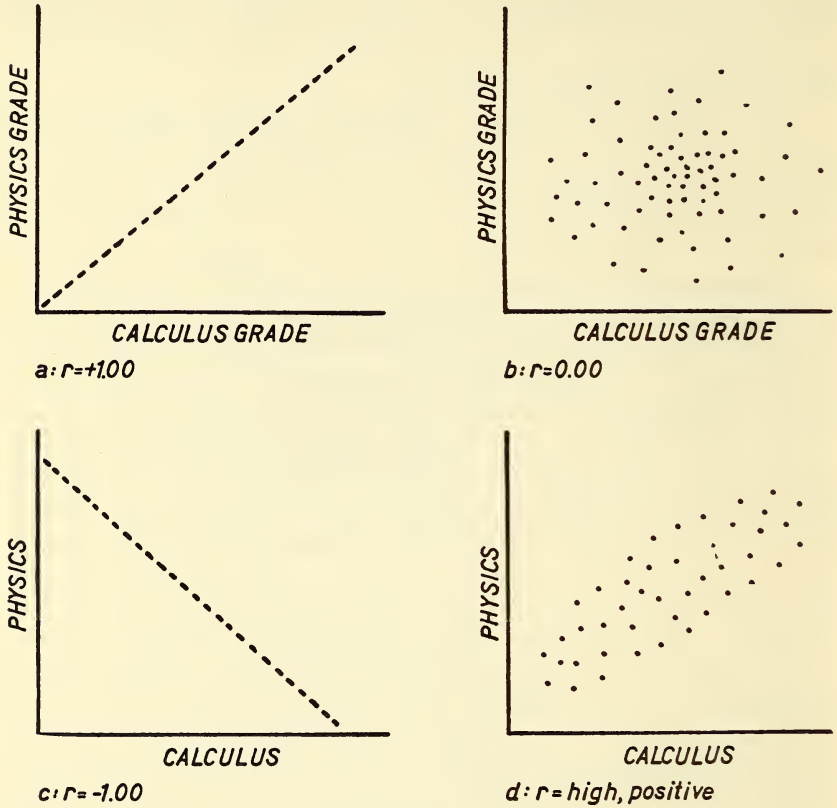


Figure 1. Scatter Diagrams Representing Varying Degrees of Correlation Between Grades in Physics and in Calculus

Each point represents the pair of grades obtained by one individual. In (*a*) and (*c*) all points fall on the dotted line.

It should be noted that the symbol r is restricted to the measure of *linear* relationships between two variables. We can write a simple equation expressing the average relationship between the two variables. Thus, we might find the following relationship for our example:

$$\text{Grade in physics} = .5 \times \text{Grade in calculus} + 50$$

The value of r will represent the accuracy of predictions made on the basis of this equation. Knowing the value of r and applying a few simple formulas, we can determine the amount of error which can be expected in using calculus grades as predictors of success in physics. The more closely r approaches 1.00, the smaller the errors which will be made in predicting.

A question which is frequently raised is, How large a value of r is needed for "good" prediction—for predictions which will be practically useful. This is not, however, a question which can be answered. The practical value of a given degree of relationship expressed by the index r varies with the circumstances. It depends upon whether we are concerned with making accurate predictions for a given individual, or with improving the average accuracy of prediction for a large group of individuals. Obviously, the closeness of relationship must be the larger in the former case than in the latter. The practical value of a relationship depends also upon the kind of prediction which is to be made. If we are to predict only whether a person will pass or fail, we do not need so large a correlation as we do for predicting whether the individual will get an A, B, or C, etc. For these reasons, and for several others, there is no absolute interpretation of a given value for r in terms of practical value in prediction. Usually, however, the writer of psychological articles shows the kind of prediction in which he is interested, and makes an interpretation of r upon this basis.

The problem of significance arises again when we deal with the coefficient of correlation. In studying the relation between two variables, we must always deal with a sample of observations. A certain degree of correlation might appear in a given sample, even though there is no relationship between the variables in the long run—on the average for all possible observations. Thus, if we were to take the records of only ten students for our example of calculus and physics, we might, by chance, find a close correlation between the two grades in this group of ten students. Ten other students, on the other hand, might show no relationship at all. Table 1 shows the amounts of correlation which may be obtained by chance with varying sizes of sample. To be significant of any general relationship, a correlation must fall outside these limits.

If, then, we were to observe a correlation of .50 in a group of twenty-five cases, we could draw no conclusions. There might be a relationship which would appear again in other groups, or there might be no such relationship at all. All we can do is to make further studies to find out. If the correlation of .50 appeared in a group of

a hundred cases, however, it would indicate that some real and dependable relationship exists. It might not be exactly .50, but we can be quite sure that *some* positive relationship exists, and so we speak of this result as a *significant* correlation.

TABLE 1

RANGE OF CORRELATIONS WHICH MAY BE OBTAINED BY CHANCE FOR VARYING SIZE OF GROUP, ASSUMING THAT NO CORRELATION EXISTS IN THE POPULATION FROM WHICH THE GROUP IS DRAWN

Number of Cases	Chance Range
5	± 1.00
10	± .94
25	± .60
50	± .42
100	± .30
1000	± .09
10000	± .03

We may also use other measures of correlation, measures which apply when there is a curved relationship between the variables, in place of the linear relationship we have discussed in showing the meaning of r, and also measures which allow us to relate several variables together. For example, we can measure the predictability of a certain variable from two or more other factors. To follow our previous example, we might be interested in predicting physics grades from both calculus grades and grades in elementary chemistry. The multiple correlation (R) is similar to the simple linear correlation in every phase of its interpretation, except that it shows the accuracy of predicting from several variables at once.

The Dependence of Variables.—The statistical measures described in the preceding section tell us whether or not some relationship between variables exists, and how accurately we can predict one variable from others. Now we wish to determine *what kind* of relationship exists, not what kind of mathematical relationship we have, but what kind of psychological or physiological dependence exists.

To accomplish this aim, we must pass beyond the simple intercorrelation of variables to some form of experimental control of other factors. To return to the example which we used throughout the section on correlation, suppose we wish to demonstrate that the physics grade is dependent upon the material which is learned in calculus, independently of the special aptitude which might show its effects in both activities. In order to do this, it is necessary to analyze the rela-

tionship between physics and calculus within groups of equal aptitude. It becomes necessary to measure this third factor of aptitude in order to control its effects upon our results. Until we are able to find out what this aptitude is, and to develop some means of measuring it, we shall be unable to solve our problem.

If we are not certain what the additional variables may be, as in this case we may not be certain what kind of aptitude could be the link between the two performances, we can resort to a mass control of a number of additional variables which are likely to include those which are critical to the problem at hand. In our example we may control intelligence, numerical ability, mechanical comprehension, and a number of other factors which are already measurable by tests, even though we do not know whether, or how, they are involved in either calculus or physics grades. The procedure would be as follows:

1. Administer all the aptitude tests to a group who had high achievement in calculus, to a group who had low achievement in calculus, and perhaps also to a group who had never taken calculus (assuming that the latter are admitted to the course in physics).
2. Equalize all three groups with respect to their aptitude test performance. A number of devices may be used to ensure that the groups are equivalent in this respect; one common method is to match each individual in one group with an individual in another group who has nearly the same aptitude test scores.
3. Compare the performance in physics with the performance in calculus for these equalized groups.

If a previous correlation in uncontrolled groups no longer appears in these matched groups, we could be sure that the predictive relationship between the two courses depends upon additional variables contained in the aptitude tests. If some correlation still remains, two interpretations are still possible. Either the knowledge of the calculus contributes directly to the physics grade, or we have failed to control some additional variable—our tests have not covered all phases of aptitude, or factors like interest and motivation have been left out of account.

Before we can draw conclusions about the dependence of variables with certainty, therefore, we must control or hold constant all potential third factors.

Let us consider another example to illustrate another way in which the same general goal may be accomplished. On page 9 we considered some statements whose accuracy we need to check before they are accepted. One statement read, "Loss of a night's sleep has a

serious effect upon performance of a job the next day." It is a statement which would probably have many endorsers among laymen. Notice that the statement is more than a statement of predictive relationships. It does not say merely that a person who had little sleep is unlikely to perform well the next day, it says that the loss of sleep is responsible for the poor performance. It is, in other words, a statement about the dependence of these two variables.

We might question first whether the predictive relationship is itself true, but let us suppose that for purposes of illustration we grant that it is correct. Suppose, in other words, that, on the average, persons with little sleep on a given night *do* perform poorly the next day on their jobs. When that has been determined, we have still no conclusive evidence that the loss of sleep is responsible. People who do not get a full night's sleep frequently have attended parties, consumed alcohol, danced, smoked more than usual, consumed unusual food, and so on. We must somehow eliminate these factors before we can determine the effect of the sleep loss proper.

Someone might suggest that these factors could be eliminated by studying only those who lost sleep because they had insomnia and "just couldn't sleep well." This procedure would evidently eliminate the factors we have mentioned so far. There are still others to be considered, however. Perhaps the person who didn't sleep well is suffering from a temporary upset in health which would also affect his work. Many persons worry about losing sleep, and *expect* the loss to have an adverse effect upon their work.

In many fields of psychological investigation, variables like the last are the most difficult to control and in some cases they have never been satisfactorily eliminated from the situation. In this respect, a laboratory study is likely to be more satisfactory than an investigation carried on directly in the shop or factory. The laboratory subjects frequently are less likely to be biased in their expectations, their other activities can be kept under observation, and their performance more carefully measured. This is not to say that an experiment in a laboratory is always to be preferred to an investigation in a more natural setting. In some instances the laboratory situation itself produces distorting variables which destroy the accuracy of the experiment, or make it difficult to apply the results to practical situations. Nevertheless there are many problems which require laboratory study in order to eliminate variables which are uncontrollable in the natural situation.

In this introductory discussion we shall not attempt to catalogue all the devices which may be used to control the effects of extraneous

variables. These examples should suffice to indicate the importance and the difficulty of the problem. As we discuss the various specific problems of efficiency, we shall become more intimately acquainted with many of these devices.

This discussion of the steps necessary, first, to establish a predictive relationship, and, secondly, to determine the interdependence of variables, should make it clear why so little faith can be placed in statements and claims based upon common sense or common observation. These common observations may be wrong even when it is a matter of simple statistical or predictive relationships. How can we be sure, from simple qualitative observation, that students who do well in calculus also do well in physics, or that a person without much sleep does not maintain his normal production the next day? We should certainly have difficulty in estimating the closeness of the relationship, and our observations would be so limited that the factor of chance could never be accurately estimated.

If we would frequently be in error about the existence of a statistical correlation between two variables when we base our opinion upon common observations, how much more opportunity for error is there when we are concerned with dependence of variables as well as predictability?

Chapter 2

THE MEANING OF EFFICIENCY OR ECONOMY OF WORK

Since the study of *efficiency* is to be the central topic of this book, and since it is the ultimate concern of almost any investigation in the field of applied psychology, it would be well to consider the meaning of the term and the implications of this meaning in some detail. In later chapters we shall show how research methods are applied to scientific and practical problems of efficiency arising in the factory, the office, and the school. In these later instances we shall find that the measurement of efficiency is only partially realized or approximated. The measurement of efficiency of the human organism at work represents a goal, or an ideal, of applied psychology which is never quite reached in practice. It is for this reason that it is especially important that we have a clear conception of the nature of this goal or ideal.

To avoid confusion, terms which are held in common by the various sciences should have similar if not identical meanings. As used in physics and mechanics, the term "efficiency" refers to the proportion of the energy input of a machine which is recovered in useful work. A similar ratio of output to input is essential if the term is to be applied to the human organism. For some purposes we can use an identical definition in physics, in physiology, and in psychology. For other purposes, however, we shall find it necessary to broaden the term somewhat beyond the strictly mechanical usage.

This broader usage of the term would make the phrase "efficient performance" equivalent to "economical performance." That is, we should consider not only immediate energy consumption, as we do when measuring mechanical efficiency, but also the over-all cost of work. The buyer of a motor car is interested not only in its fuel consumption (the mechanical efficiency of the engine) but in the cost per mile of operation, including initial expense, depreciation, and repairs, along with the fuel. Similarly, for practical psychological problems, it may not be enough to know the mechanical efficiency of the body in a given task. Other elements of cost may be much more important to economical operation.

Note that even though we are not using the term efficiency in its exact physical meaning, we are, nevertheless, retaining the important feature of a ratio between output and input. We have simply broadened the concept of input beyond the energy immediately consumed. The other elements of cost also ultimately involve energy, but it is not convenient or possible to express them all in energy terms.

Misuses of the Term "Efficiency."—In common speech, the term "efficiency" is used in many vague and often misleading senses. Some of these misuses of the term have led to misunderstandings and errors. The commonest confusion is between efficiency and rate of performance. Even some psychological texts have a chapter entitled "efficiency" which begins with a careful definition in terms of output and input, but confuses efficiency with rate of performance before the chapter is over.

If the rate of performance is increased, efficiency may go up or down, or it may be unaffected, depending upon simultaneous changes in the cost of work. If we do not know anything about the cost of work to the individual, we are safer and more honest if we simply report the facts in terms of rate of performance, and avoid using the term efficiency.

Input, Cost, Energy, and Effort.—In addition to the energy transformation directly measurable during performance (metabolism), the input or cost of work includes such factors as fatigue, the time of the individual, effects upon his health, his personal adjustment in society, his ability to enjoy leisure time and recreation, and possibly other factors. In most instances, these factors are more important than energy. We need a convenient term to cover this varied list of items. The term *energy* is frequently used in common speech to refer to various aspects of the cost of work, but this is likely to lead to confusion with the strictly physical terminology involved in physiological studies of metabolism, as will be seen later. The term *input* is often closely identified with fuel consumption, and *cost* might imply a monetary basis of evaluation. As applied to the human organism, these implications should be avoided; they will be avoided in this book. *Input* and *cost* will therefore be used as equivalent terms, with a very general meaning. Included in input will be *any* adverse effect of the work, anything which the individual expends or uses up in carrying on the activity under study.

Input involves many different factors or elements which will be considered in greater detail later. In order to avoid confusion, however, we may note how two of the important elements of input will be

defined in this book. *Energy* expenditure will be used to designate only those aspects of activity which can be expressed in terms of physical units—calories, watts, etc. It will refer only to that phase of input which is taken into account in the physical meaning of the term efficiency. *Effort* will be defined in a way which is somewhat different from common usage. This is necessary because the term, while used constantly in common speech, has no clear meaning in that usage which would permit exact study or measurement. The most useful definition of *effort* would be that it describes *the relationship between actual rate of performance and the capacity of the individual at a given time.* Effort involves energy, but it also involves other things, so that the two terms should not be equated or confused. For example, there are many tasks of the organism which involve very little expenditure of energy, but which may be carried on at a very high level of effort; (for example, mental arithmetic). This definition of effort includes many of the usages of common speech, but it makes the meaning more exact and shows us what is required if we are to measure effort.[1] It will also appear later that this conception of effort provides us with a key to the study of the relationship between motivation and efficiency.

The Importance of Measuring Efficiency.—Some may object at this point that any supervisor, industrial manager, or even a teacher is mainly concerned with output and results rather than with efficiency. They may also say, as they discover some of the difficulties of studying efficiency, that the notion complicates matters too much; that everything would be simpler if we just forgot about it. Problems are not solved by ignoring them, however, and some of the problems we should be ignoring would be extremely important.

Let us consider some of the places in which an adequate analysis of efficiency is required. In some cases the reader will be surprised to find wide ramifications of this problem.

1. The problem of wage and time standards in industry. The ideal here would be to establish payment methods so that each task would be paid for in proportion to the *effort* required to do it. As we shall see when we discuss *time study* and related problems, this is essentially what various methods now in use are designed to do. They are little more than rules of thumb, however, and will remain so until our studies of effort and the cost of work have progressed further.

[1] This meaning of the term has been used explicitly in a recent investigation of college aptitude testing, with results which show considerable promise. See P. R. Sappenfield, Prediction of college scholarship for groups having effort indices of restricted range, *Journal of Applied Psychology,* 1943, *27,* 448-451.

The relationship can be seen best in a simple example. Two men are working at jobs requiring similar training and schooling, and involving about equal risk or danger. In other words, their "investment" is about the same. One man produces 100 pieces of the part he is working on. The other has a different task at which he is able to produce 50 pieces per hour. What should be their relative pay? If we wish to use a system of rewards for hard work, we must raise the question of effort. Does it take more effort to produce 100 pieces on the first job in comparison to the 50 pieces on the second job? Various crude solutions of the problem are now in use, but their validity is questionable because they have no satisfactory way of estimating effort. (See Chapter 10.)

2. Methods of increasing productivity of a given department, plant, or industry. If inadequate consideration is given to the input side of worker performance, a method of work which at first appears to raise production may eventually reduce it. This is due to the possibility that some of the cost does not show its effect for some time. A method which is to be permanently effective must consider these long-term elements of cost and try to find out about them in advance. Otherwise, the only way they will be discovered will be through an expensive process of long-term trial and error.

3. Hidden expenses. The overhead of an industrial enterprise is definitely affected by turnover, accident rates, discontent, and complaints. These, in turn, may be the result of methods which lead to high productivity, but relatively low efficiency of human performance. In other words, inefficient performance of the worker is, in the last analysis, a condition of inefficient enterprise even though this may not appear at once under superficial examination.

Applying the Concept of Efficiency: An Illustrative Example.— In order to make the notion of efficiency more concrete, let us consider a single problem which occurs in modern industry. The exact methods of solving this problem must await later consideration, when we have studied in more detail the various experimental methods of measuring the cost of work to the individual. At present we shall merely examine the implications of the problem from the point of view of efficiency as the term is used here.

As our example, we chose the general problem of specialization and its effect upon the efficiency of the worker. Of course, specialization is likely to have different effects for different conditions and types of work, so that we shall have to make the problem more specific by considering its effect upon a single task and comparing two

specific levels of specialization. Here is a task consisting of light, manual, semi-skilled work—assembling small parts of a typewriter or adding machine. A certain mechanism consists of two main subdivisions, each made up of six individual parts. The two subdivisions are assembled separately, then joined together to form the completed mechanism. At present each worker performs the whole task, assembling one subdivision, then the other, and next completing the final assembly of the mechanism.

Since there are a number of workers performing this same task, it is proposed that they be specialized, with one worker assembling only subdivision A, another assembling subdivision B, and a third worker completing the mechanism. The problem before us is how we may determine the relative efficiency of the newly proposed method as compared with the former procedure.

The practical industrial supervisor, and probably most readers, would feel that they would want to see the job in operation, to know what the parts are like, what tools are used, and all of the details of the job. Naturally, all details should be before us if we are to work out the answer to our question. But we must stress the fact that we cannot answer our question adequately by applying "common sense" to the knowledge of the nature of the work. To attempt to do so would scarcely be sound scientific procedure, although it might be intelligent guessing. Yet, in practice, many problems of this sort are solved merely by inspecting the job and stating, "Obviously, specialization is better for this job than the present method."

We shall deliberately refrain, therefore, from giving any more detailed information about the nature of the work involved in our problem. We do not need these details in order to consider the general methodology which would be applied to a sound solution of the problem. No matter how familiar you are with the detailed operations of the job, there is no equivalent substitute for an experimental study of both methods which would determine, in as direct a manner as possible, the relative efficiency of the two methods.

For this kind of study it would be necessary that trials of both methods be carried on by workers of equivalent skill, with the same incentives, over a considerable period of time. This has frequently been done, but in most cases the experimenters have been satisfied to determine the relative *productivity* of the two methods. Suppose we find that the unspecialized method produces an average of 125 pieces per day, while the specialized method yields 150 pieces per day, and that the difference is statistically significant. What have we accomplished so far?

Such a result is important, but it is only half the story. We know that the numerator of the fraction defining efficiency has increased, but what of the denominator? Until we know that, we have no way of determining whether the efficiency of the specialized method is higher or lower than that for the former method. To assume that the specialized method is "better" is to ignore the denominator completely.

What must be considered in order to complete the analysis of the relative economy of the two methods of performing the task? In common terminology, we wish to know whether the individual worker is exerting himself more under the new regime, whether he is more fatigued at the end of the day, whether there is any accumulation of fatigue over a period of days or weeks, whether the new method makes him "jittery" and has a harmful effect upon his appetite, his enjoyment of leisure, and his general health. As we have pointed out above, to consider these factors is not a matter of pure altruism on the part of management. Increases in these factors lower the "morale" of the workers, make supervision more difficult, increase the turnover and absentee rates, and generally affect the overhead of operation. Unless these elements are taken into account, the balance sheet for the job gives an entirely erroneous impression of the economy of the operation from the point of view of management, as well as of the individual performing the task.

It is possible that these potential losses are partially offset by certain gains on the part of the worker. He may get increased satisfaction from the work if his rate of output is higher. His increased pay may be important enough to him to compensate for the additional cost. Satisfaction from the work may be considered positively as an addition to the output beyond the number of assemblies produced, or it may be considered negatively as dissatisfaction and added to the cost. For reasons which will come out later, we shall place satisfaction and interest in the numerator of our efficiency ratio in place of adding dissatisfaction to the total cost. Numerically the result is the same in either case.

These, then, are the general elements which must be considered along with production rate to form an over-all picture of the relative efficiency of the two methods under consideration. A statement of general terms is not enough, however, to permit us to compare efficiencies of different methods or conditions. We must be able to *measure* these changing factors in cost. The output in our example has increased 20 per cent as a result of the change of method. If fatigue, loss of health, or effort has increased by only 10 per cent, the

new method is more efficient than the old. The efficiency is not 20 per cent greater, but at least it has increased, and the employer is interested in the additional output. Suppose, however, that the cost of the work has increased 20 per cent or more. Then it is doubtful whether the employer would gain in the long run by the use of the new method, even though the productivity has increased. With this increased cost of work, it is even doubtful that the productivity would remain at this high level over a long period of time.

In the preceding paragraphs we have talked glibly and rather loosely about what should be done to analyze the efficiency of a method or condition of work. By our use of such easily understood words as "fatigue," "health," and "satisfaction," we may have emphasized the complete force of the term efficiency at the expense of making the problem appear simple. If this is the case, a second thought would show that the accurate study of efficiency is an extensive and difficult task. The first necessary step is fundamental research devoted to developing methods of *measuring* these factors in the cost of work. Only after some progress of this kind is made will it be possible to analyze efficiency in practical problems of the kind we have taken for our present illustrative example.

Our first task in succeeding chapters of this book will be to evaluate progress which has been made along this general line of research, to consider what techniques are of some use in practical research problems, and what further fundamental research needs to be carried out in order to increase progress in this field.

It is the difficulty of the task of evaluating the cost of work which is the principal justification for considering only output in comparing methods of doing work in industry. Where this must be done, however, it is dangerous to confuse the issue by equating measurement of output with measurement of efficiency. If the industrial manager was unable or unwilling to measure relative efficiency, he should say so, and speak of the relative *productivity* instead.

Rate of Work, Efficiency in Emergencies, and the Aims of the Study of Efficiency.—Rate of work is obviously one of the factors which affects efficiency of performance. In many activities it has been found that a moderate pace is more economical than either a very slow or a very rapid speed of work. If the individual works very slowly, his "overhead" is high in proportion to what he accomplishes, since he must go on living, standing, walking about. Also, he may use more awkward rhythms of movement. If his speed of work approaches the maximum rate of which the worker is capable, he ap-

parently adopts again relatively wasteful methods of work, just as an industry is likely to waste more in periods of peak production. Thus, output does not go up in proportion to the effort expended.

If this relationship between rate of work and efficiency can be regarded as a generalization (as it apparently can), it raises a question about the importance of the study of efficiency under emergency conditions. Those who read a discussion of efficiency of the individual worker during wartime were likely to feel that this is a topic worth considering in peacetime, but that it had little relevance to wartime emergencies. One cannot achieve maximal efficiency at high speeds of work; therefore they concluded that we must await the more moderate pace of peacetime before taking an interest in the problem.

The determination of optimal efficiency is always, however, a relative matter. Under any given conditions, certain factors cannot be changed. Our problem is to adjust to the most satisfactory level those factors which are under control. During war we do not have complete liberty to choose the rate of work which leads to maximum output per unit of cost. The problem is, rather, to determine conditions of efficiency which can be controlled and which improve efficiency at a *high* rate of performance. Variations in speed still need to be considered, since the speed could increase to the point where errors, accidents, and disruptions in production would be so frequent that the *net* output would be reduced.

Another way of looking at this same problem might also be enlightening, even though it is largely speculative at present. In emergency conditions, the worker is willing to devote more effort to his work than he is in normal times. Nevertheless there are limits to the total input which is available, especially if the "emergency" is long lasting. Under these conditions, the problem of efficiency is to find out how this limited input can be used to greatest advantage. Rate of work is a variable within certain limits because the worker might use his quota of input in six hours per day or spread it over a longer working day. The only limitations upon rate of work are that the worker is expected to use his quota completely, and his rate of expenditure must not be so great as to introduce errors and accidents. In other words, rate of work is still a variable factor in efficiency, even under these more limited conditions where we cannot aim for the absolutely optimum speed of performance.

It is also conceivable that the optimal rate of performance might not be put into practice even in peacetime, for a number of reasons. The optimal rate for a single individual might be so slow that it would produce a labor shortage and an inadequate supply of goods. Even

from the individual's point of view, the optimal rate might not be satisfactory. By optimal rate we mean that rate at which there is a maximum output per unit of input. Suppose, however, that this leads to a very low total cost to the individual per day. Perhaps even after we have allowed for the effort he can expend upon his hobbies, amusements, home affairs, and the like, the worker is still left with a considerable reserve which is not used at all.

The problem of rate of work is, then, very complex at all times. The discussion in the preceding paragraphs was a statement of various possible relationships rather than of established facts. The mere existence of these possibilities should indicate that a more careful analysis of over-all efficiency is needed, both in wartime and during peace. Conjectures and opinions are made the basis of many arguments about problems of fair wages and hours of work, restriction of output, "featherbedding," and the like. These arguments must remain in the realm of conjecture until more adequate research upon efficiency can be related to these problems.

Applications of the results of research will not always be directed toward absolutely optimal conditions for all factors. There would be no point in allowing effort to go to waste in achieving optimal efficiency. Other factors, however, will not only raise efficiency at any given rate of performance, but they will also permit a higher output for any given expenditure per day. Whether we seek an optimum condition, or a condition which merely avoids extreme *in*efficiency, the general concept of efficiency is fundamental in our thinking.

Poffenberger has stated an "ideal of human efficiency" which "would be the production of the maximal output of the highest quality in the shortest time, with the least expenditure of energy and with the maximum satisfaction." [2] This statement is, in the opinion of the present writer, incomplete in several ways. Only energy and time are considered as elements of input, although Poffenberger shows in the remainder of his discussion that he intends to include fatigue and other elements of cost as well. Apart from this condensation of the statement, there is a more serious drawback. A maximum of output and a minimum of cost are incompatible. Instead, we must seek a compromise between these two distinct aims. A better statement might be: "the greatest output and satisfaction which can be attained without *excessive* cost, understanding by excessive cost a level of effort which leads to accumulation of fatigue, inability to

---

[2] A. T. Poffenberger, *Principles of Applied Psychology,* New York, D. Appleton-Century Co., Inc., 1942, p. 364.

enjoy leisure, loss of health, premature aging, increased liability to accidents."

This revised statement of the broad aims of research upon efficiency should, we believe, answer a critic of Poffenberger who raises the question "Why be efficient?" [3] This critic, Johnson, believes that many of us are *too* efficient, that those who grow fat might be better off to use more energy in their work, rather than less. Johnson's notion of efficiency is restricted to mechanical efficiency; it ignores the possibility that a performance which has a low cost in terms of energy may have a high cost in terms of other elements included in our own broader meaning of efficiency. Even if we restrict attention to the mechanical side of efficiency, the choice of a method of work need not be dictated solely by an optimal efficiency ratio, as the preceding discussions should indicate.[4]

Total Efficiency. There is another stumbling block in the measurement of total cost of work beyond that of the number and complexity of factors included. This difficulty is that of summing up the elements of cost. Obviously the cost of human work to the individual cannot be placed on a dollars and cents basis, even very roughly. It is equally evident that many of the elements, such as discontent or loss of health, cannot be expressed in energy terms. How, then, is it possible to make any useful research approach to the problem of efficiency? It can only be done piecemeal and indirectly. If we are able to hold most items of cost at a nearly constant level while one item varies over a wide range, we can draw conclusions about efficiency from this one factor alone. If a given factor reduces all or most elements of cost at the same time, again it is easy to draw our conclusions.

When, however, we find divergent effects upon the various items of cost, the problem is one which can be solved only in terms of some theory of value. Suppose, for example, that a change in working conditions increased fatigue but decreased discontent. The decision as to the effect upon efficiency or the desirability of the conditions can be made only by coming to a decision about the relative importance of fatigue and discontent to the individual, and in the long run. Perhaps such a question could eventually be answered by research.

---

[3] See H. M. Johnson's review of Poffenberger's *Principles of Applied Psychology, American Journal of Psychology,* 1943, *56,* 613-619.

[4] Parenthetically, we might also raise the question of whether growing fat is necessarily an indicator of high efficiency, even of the mechanical kind. Perhaps it indicates the efficiency of the digestive and glandular processes more exactly than it does the efficiency of the activities involved in work.

This is an instance, like others which we shall meet later, where the applied psychologist has to take a somewhat different approach from the psychologist who is interested in "pure science." The latter, particularly one who calls himself an "operationist," would say that the problem of efficiency, as we have formulated it, is not a scientific problem. What do we gain, he would say, by using a term which refers to elements which cannot be measured or satisfactorily combined? He would prefer to take one aspect of cost which is readily open to observation and measurement, and define the cost of work in these terms. Then, when he uses the term *efficiency,* he will refer to a perfectly definite, measurable thing. Suppose, for example, that we define the cost of work in terms of the oxygen consumed per unit of work, which is the common way of measuring the mechanical efficiency of muscular work. We would decide that the term *efficiency* will mean nothing more nor less than the oxygen consumpion per unit of work.

The trouble with this approach, from the point of view of the applied psychologist, is that it applies only to work where physical energy expenditure is great. In other tasks, other elements of cost are more important, and we cannot diminish their importance by refusing to consider them. It is better, therefore, frankly to face the fact that some of the elements of cost which would be essential to a complete solution of the problem of efficiency are not amenable to available methods. It makes us realize the limitations of conclusions based upon the measurable elements, and shows us the direction which research must take in order to reach more complete solutions of the problem.

Interrelationships of Factors in Efficiency.—The various factors in efficiency are interlocking in a variety of ways. Some of the interrelationships have been discussed at various stages of the development in this chapter. It might be well, however, to bring together all these interdependencies to form an integrated pattern, and to bring out more clearly the meaning of the various terms which we have used in discussing efficiency.

A worker is performing a task under specified conditions and with a specified method. At a given moment, and under these specified conditions, the worker has an upper limit of speed of work beyond which he cannot go. It is this momentary, conditional limit which we shall speak of as his *capacity.*[5] *Effort* is a term describing the rate of

[5] It will be noted that this notion of capacity is distinct from the meaning usually implied in psychological texts. Usually, capacity is used interchangeably with *aptitude,* or potentiality to learn a given performance. In the context of efficiency,

work in relation to capacity. This measure of effort can thus be changed by either a change in rate of work or a change in capacity. If, for example, the capacity of the worker declines while his absolute rate of work remains constant, his effort has increased. Conversely, rate of work can increase without any increase in effort. If a new method of doing the job is introduced, a method which makes the work easier and raises the capacity of the worker, the rate of work can be increased proportionately, to leave the level of effort at its original value.

Other phases of the cost of work are, in turn, functions of effort. Thus we shall assume that fatigue, energy expenditure, loss of health are each dependent upon the degree of effort involved in an activity. The exact relationships are not known. It is likely that a different function describes the relationship of each of these factors to effort. Each is probably also dependent upon other things besides effort in the task itself, but our main concern is with the relationship to effort. It is also very probable that the functions are not linear. That is, fatigue is not proportional to effort, but probably increases disproportionately when effort approaches the maximum. Fatigue may also increase when effort goes down to a low level. It is also likely that these relationships vary with the kind of work. Energy expenditure, for example, may bear a close relationship to effort in heavy muscular endeavors, but show little effect of other kinds of effort.

It is apparent that a single research project cannot take all of these potential relationships into account. Ordinarily it is necessary to limit the study to the observation of a few factors. Thus, energy expenditure may be the only element of cost which can be accurately observed; in other instances attention must be concentrated upon fatigue. It is seldom that effort is taken into account directly and related to fatigue or other elements of cost. Instead, fatigue is related directly to some external factor such as noise, or to rate of work as a substitute for a measure of effort.

In spite of these necessary shortcomings in research procedures, all the interrelationships just discussed should be kept in mind in interpreting the results of specific experiments. Otherwise there is a serious possibility of error or misconception of the meaning of the results.

---

however, the meaning given in this text is usually adopted, separating the concepts of capacity and of aptitude. Capacity is thus a limit which is never reached in practice, and is a concept closely similar to the concept of *limit* in mathematics. Even though this limit is never reached, and therefore capacity cannot be directly measured, it can be estimated indirectly, and is a construct which is useful in our thinking about human activity, just as the concept of limit is useful in mathematics.

Summary of the Concept of Efficiency.—The term *efficiency* involves many complex factors. In this book it will designate a ratio between output and input (cost of work or effort). To be complete, the output term should include not only the "objective," tangible products of the work, but also the pleasure and satisfaction which the worker may derive from his accomplishment. The input factor of the ratio involves a complex set of factors, since we wish to include all elements of the cost of work to the worker.

It would be pleasant to simplify the concept of efficiency by reducing the number of terms which are to be taken into consideration, and by simplifying the elements themselves. Such a simplification would be a loss to applied psychology, however, since this is a field in which we must accept our problems from everyday living, and do our best to cope with them as they appear. We cannot avoid these problems forever, even though we may not feel able to cope with them in their complete form at the moment. To retain the complex meaning of the term efficiency is to lay out a general program for applied psychology which is to be solved little by little. The purpose of the coming chapters will be to survey what has been done in this program. To ignore the parts of the program which are still incomplete, would, however, give the reader an utterly false notion of progress in applied psychology.

There can be no doubt that progress in this program will have marked practical effects. Some of our most important problems of personnel and labor relations, such as fair wage rates, hours of work, work methods, morale, and output restriction, depend in their last analysis upon the facts of efficiency for their solution.

In this discussion of the implications of the term *efficiency,* we have, perhaps, stressed the study of *input* too much in relation to the study of output. The reason for this is that we wished to emphasize those phases of the problem which are most commonly neglected or minimized. We do not intend, however, to go to the other extreme and to neglect the very important study of productivity. Output is an important subject for study in its own right, as well as an integral part of the study of efficiency. If the effect of a certain factor upon output, but not upon efficiency, is known, this information can be useful and we shall not intentionally slight it.

Chapter 3

MEASURING THE COST OF WORK: I. METABOLISM AND MUSCULAR FATIGUE

The measurement of output is relatively easy—at least it is easy enough so that we can take that phase of efficiency measurement for granted until we have become more familiar with the problems which we face in dealing with the cost of work. The purpose of this chapter, and of the next, will be to familiarize the reader with the general techniques for studying and evaluating the cost or input. We shall consider the techniques of measuring the cost of work according to the class of activity for which each technique finds its principal use.

It is common to divide work into two main classes, "mental" and "muscular" (or "physical"). Such a separation originally implied a separation of mind and body which scientists no longer regard as a valid or useful separation. Research studies of the cost of work still fall roughly into two general classes, however, even though the original basis of the distinction between "mental" and "muscular" work has been lost, and the adjectives no longer have their original meaning.

We shall include under muscular work only those kinds of activity in which the *primary* aim is the development of mechanical *force*. Muscular work would include tasks which make heavy demands upon the whole muscular system—running, walking, carrying large loads, shoveling, scrubbing, and the like. It would also include those tasks which involve the development of force by a restricted muscle group, such as lifting a weight with one arm or one finger while the rest of the body is supported.

Perhaps it would be better to speak of "sedentary" work rather than "mental" work as our second main division of work. The reason for preferring such a terminology would be to avoid the unfortunate implication that "mental" work involves only the "mind," which in popular usage frequently implies the brain alone. Nevertheless we shall, for the sake of brevity, use the term "mental" with the warning that it means only that we are dealing with a kind of work in which control, timing, skill, and direction of activity are more im-

portant than muscular force, whether or not muscular activity is involved. Since most occupations of the sedentary kind fall into this classification, it seems that "sedentary" would be as good an adjective as "mental" although it is a less familiar usage.

There are many activities which are difficult to classify into either of these main divisions. All would agree that loading a wagon with hay belongs to the muscular class of work. Solving a problem in the calculus evidently belongs to the sedentary or mental classification. What about such a task as typing, however? Such a task involves some of the characteristics of each class. Muscular force is responsible for the impression on the page, but the differentiation between adequate and inadequate performance is primarily in terms of accuracy and pattern of movements rather than force. For some purposes, therefore, we could study typing as a kind of muscular work, but we should be considering only one aspect of the task. In the over-all view of the task, typing should probably be classed with mental work. This example should emphasize the fact that the meaning of *muscular work,* as used in this discussion, is more restricted than it is in common usage. Some tasks which we think of as manual jobs may still not be called muscular work because *force* is not a primary feature of the task.

A classification of kinds of work should probably be made eventually on the basis of the internal bodily activities involved, and in terms of the kind of effect which that work develops in the body. The present classification is partly in terms of the nature of the outcome of the work, as well as conjecture as to the bodily activities. We need some kind of preliminary classification, however, until a more adequate one can be devised; otherwise it is difficult to organize the material upon the cost of work now available.

In the following discussion of muscular work, we shall find that several special divisions of the class are beginning to emerge. The effects of fatigue in these classes appear to be quite distinct. It will therefore be useful to keep these subdivisions in view as we go along. The divisions are listed in Table 2.

Much more is known of the bodily effects of some of these types of work than of others. For example, more research has been directed toward the first type of "dromal" task than to the steady grind, probably because the first lends itself more readily to laboratory investigation, although the second is of greater practical significance.

The central interest of this book is not in muscular work, although the present chapter is primarily concerned with the topic. Our purpose is to introduce the reader to the techniques of measuring cost of

work which have been developed in this field, and to do this without presupposing much background in physiology. This survey is essential background for studying the cost of sedentary or mental work. As psychologists, we are more concerned with mental work, and it occupies a more important place in business, industry, and education. Nevertheless, research upon muscular work has an interest and importance of its own, as well as furnishing a needed background for other studies.

TABLE 2

TYPES OF MUSCULAR WORK *

1. Rapid expenditure of a large amount of energy ("dromal" tasks). *E.g.,* running a competitive race.
2. Steady grind. *E.g.,* a long day, walking under a load, stoking, etc.
3. Repetitive local task with high local energy expenditure. *E.g.,* lifting a weight with the finger.
4. Postural restrictions ("static work"). *E.g.,* maintaining an awkward position; holding a weight in outstretched hand.

* Modified from T. A. Ryan, Varieties of fatigue, *American Journal of Psychology,* 1944, 57, 565-569.

Energy Cost of Work (Fuel Consumption)

In discussing muscular force, we naturally consider first its cost in terms of the energy (in calories or B.T.U.) released in the body—energy which is ultimately derived from food. Because the body has so many provisions for storage of energy-producing materials, food intake can give us only a very rough indication of the energy used at any given time, even after we allow for waste through excretion. Oxygen consumption during any interval of time is, however, proportional to the energy released in the body, since oxygen cannot be stored in quantities large enough to affect the results. To be sure, the energy equivalent of a given volume of oxygen varies somewhat with the nature of the food which is being oxidized. It is possible to estimate the proportion of various nutrients being used and make allowances for this variation, but with an average diet the variation is not great enough to affect the results in any important way. One liter of oxygen corresponds to about 5 (large) calories (Cal.) of energy liberated (4.686 Cal. for pure fat to 5.047 Cal. for pure carbohydrate).[1]

[1] F. G. Benedict and H. Murschhauser, Energy transformations during horizontal walking, Carnegie Institution of Washington, 1915, Publication No. 231, p. 8.

This correspondence between oxygen consumption and the energy released in the body holds over a period of time, but it is inexact for any short span of time, such as a few minutes. If a man engages in heavy muscular effort for a short time, we find that his oxygen consumption is high not only during the time when he is working, but also for a considerable period after he has stopped work altogether. During work, he contracts an *oxygen debt* which is repaid during the following rest period.

The explanation for the appearance of the oxygen debt is that the oxygen is not used directly in the chemical processes of muscular contraction themselves. It is known that these processes can go on in the absence of oxygen. The muscles soon lose their power to contract, however, unless oxygen is available. Thus the oxygen is necessary for rebuilding the contractility of the muscle, but it is not necessary that all this recovery process be accomplished during the work itself. In fact, experiments indicate that a certain amount of lag is beneficial, and that the muscles operate more effectively after a certain oxygen debt has been built up.

In order to determine the total energy cost of a certain muscular task, therefore, it is necessary to find the total amount of oxygen which is required to bring the muscles back to their original state before the work began. This means that the oxygen debt must be included in the total oxygen consumption. We must measure the extra oxygen used during work, and add to that the *extra* oxygen consumed during the rest period after work. It is especially important to take the oxygen debt into account for short sprints or other activities involving a rapid expenditure of energy.

In studies of work, such as lifting weights, the energy involved in doing the work is separated from the energy required to maintain the body in its resting state. Suppose, for example, that a man uses an average of .25 liter of oxygen per minute while he is resting. While he is lifting a given weight in a prescribed manner, he averages .45 liter of oxygen per minute, (including the "oxygen debt"). The difference represents the amount which is required to supply energy *for the work*. Multiplying the difference by 5 gives us the energy in calories which is required to perform the task for one minute, (in this example, 1 Cal.). (See summary of computations in Table 3 on page 37.)

Let us suppose that the work consisted in lifting a 25-pound weight, at the rate of four lifts per minute through a distance of three feet. He has thus accomplished 300 foot-pounds of work, and has used one calorie in doing so. If we had a perfect machine in which all energy

went into useful work, the amount of energy this man has consumed would have been converted into 3085 foot-pounds of work. Therefore the mechanical efficiency of this performance is about 10 per cent.

TABLE 3

ROUGH COMPUTATION OF MECHANICAL EFFICIENCY

Input:	
Total oxygen used during work (including oxygen debt)	.45 liters per min.
Oxygen consumption during rest	.25 " " "
Net oxygen consumption (increase due to work)	.20 " " "
Conversion factor (1 liter = 5 Cal.)	× 5
Net input (energy consumption due to work)	1.00 Cal. " "
Work equivalent of input (1 Cal. = 3085 F. P.)	3085 F. P.
Output (external work done)	300 " " per min.

$$\text{Mechanical efficiency (of muscular system used)} = 100 \times \frac{\text{Physical Work}}{\text{Work equivalent of input}}$$

$$= 100 \times \frac{300}{3085} = 10 \text{ per cent}$$

The way in which this method could be used for determining the value of different methods of work is obvious. Suppose, for example, it is decided to double the load and cut the rate in half. Then the work per minute would be the same as before. If, now, it is found that the oxygen consumption is more than it was before, the new procedure is clearly less efficient.

In laboratory studies it is easy to choose tasks which permit ready computation of the work done in standard physical units such as foot-pounds. The results of these experiments are suggestive and useful in many ways, but their application to different tasks with different patterns of muscular movement is very doubtful. Fortunately, many tasks can be studied by the metabolic method, even though their output cannot be expressed in a standard set of units. Horizontal walking, for example, could be converted into equivalent work units only by a complex physical study of the forces involved.

It so happens that this difficulty is not a very serious one. We very seldom desire to compare the efficiency of performance of two dissimilar tasks, such as lifting weights and walking. If we are comparing different conditions of walking, or different rates of walking, we can simply use any units of output which are convenient—distance travelled, or distance times the weight of the body and its load. We are not able to speak of "per cent efficiency" in these circumstances, since input and output are not in comparable physical units. We secure an index proportional to the efficiency, however, and we are interested only in relative efficiency anyway. In some cases we can hold the output constant and observe the metabolic cost alone as the indicator of efficiency.

Applications to Muscular Work.—Table 4 shows the results of one study of efficiency in a muscular task in which metabolic measures are employed. Atzler shows here the relation of height of a crank axis and the load to energy consumption per unit of work. In other phases of his experiment he also studied the effect of the rate of turning and the radius of the crank. The latter factors were held constant in the experiments which are reported in Table 4.

TABLE 4

Cost of Work in Lifting Loads by Crank*

(Calories per meter-kilogram)

Height of Crank Axis in cm.	Load (meter-kilograms per turn) 6.5	13.0	19.5	26.0	32.5
55.3	27.8	16.5	13.8	14.4	16.5
82.7	26.1	15.5	13.9	13.5	14.6
114.3	17.0	13.5	12.4	12.1	14.2
162.2	22.2	18.4	18.5	20.5	22.2

* E. Atzler, Probleme und Aufgaben der Arbeitsphysiologie, *Ergebnisse der Physiologie,* 1928, *27,* 720. This article (pp. 709-779) gives a number of other examples of this same method. Other examples are to be found in another article by the same author, Arbeitsphysiologie: Zweiter Teil, *Ergebnisse der Physiologie,* 1939, *41,* 164-291.

These results are a good instance of the manner in which experiments upon metabolic cost could be used in deciding questions of the design of machines and tools. It will be noted that greatest efficiency is obtained when the crank axis is 114 cm. from the floor, regardless of the load. That is, the energy (calories) per unit of work (meter-kilogram) is at a minimum for this particular height. If the height is increased or reduced, the efficiency is reduced.

It will also be seen that the very light load and the heaviest load are both less efficiently handled than the intermediate loads. The optimum load varied, however, with changes in the height of the axis.

TABLE 5

Relative Energy Cost of Pulling vs. Pushing *

(Calories per meter-kilogram)

Weight	Pulling (Two hands, grip 100 cm. from ground)	Pushing (Grip 75 cm. from ground)
10.27	10.67 Cal./mkg.	9.16
11.64	10.21	9.04
13.56	9.76	8.72
16.06	9.90	8.96

* Atzler, *op. cit. supra,* p. 742.

This set of data also illustrates a common inadequacy of experiments upon energy cost. This is the tendency to ignore individual variation and chance variation, which is understandable when we consider the great number of observations which must be made on each subject. Atzler's study involved only two subjects. They were, however, quite different in build and weight, and their results were closely similar. For this reason Atzler feels that a study of a wider range of individuals is unnecessary. Perhaps this is justified in this case, but the same tendency to work with very few subjects is often found in studies where there is much more likely to be significant variation from individual to individual.

We are not told how many observations were made to establish each of the values of the table (another common tendency among some physiological investigators). For this reason we cannot say whether some of the slight variations reported there are more than might be expected by chance.

Another kind of work is analyzed in Table 5, also by Atzler. This table represents the cost of work done in drawing or pushing carts with varying loads. It will be seen that pushing is more efficient than pulling the load for each of the weights studied. It will also be noted that, within the range of this experiment, the efficiency increases as the weight of the load increases. In another part of the study it was found that the handgrip for pulling the load was most efficiently placed at 100 cm. from the ground. A variation of 15 cm. in either direction increased the energy consumption per unit of work.

A third example is based upon a large number of studies of walking and running. Table 6 shows results assembled by Poffenberger to indicate the efficiency of varied rates of locomotion.

TABLE 6*

ENERGY EXPENDED IN HUMAN LOCOMOTION

Activity	c.c. of O_2 per Minute	c.c. of O_2 per Mile
Lying down in post-absorptive state	237	.....
Standing still	328	.....
Walking 1.66 m.p.h.	606	21816
2.00 m.p.h.	668	20040
3.00 m.p.h.	907	18140
4.00 m.p.h.	1182	17730
5.00 m.p.h.	2125	25500
Running 6.50 m.p.h.	2868	26472
7.50 m.p.h.	3395	27160
9.00 m.p.h.	3834	25534
10.00 m.p.h.	3910	23460

* Reprinted from A. T. Poffenberger, *Principles of Applied Psychology*, New York, D. Appleton-Century Co., Inc., Copyright, 1942, p. 117.

This example brings up the question of how much of the oxygen consumption to include in the cost of work. The "c.c. of O_2 per mile" listed in Poffenberger's table is based upon the *total* oxygen consumption, including not only the oxygen employed in the muscular activities of locomotion, but also oxygen which is used for the basal metabolism. He is concerned, therefore, with something more than we should include in the strict meaning of "mechanical efficiency." Some of the *overhead* is included as well, in the sense that oxygen used in general maintenance of the body is also counted. If we were dealing with different methods of doing a job at the same fixed rate, overhead would remain constant and it would make little difference whether it were included or not. When the rate of work varies, the overhead *per unit of work* (per mile, in this instance) varies. That is, the 237 c.c. resting consumption is used sixty times in a mile at the rate of one mile per hour, but only six times in a mile at the ten-mile rate.

The problem of whether or not to include the overhead is to be decided by the purpose of a study. If a physiologist wishes to find the mechanical efficiency of a set of muscles used in a certain task, he should use only the excess of oxygen over resting values in computing input. If we are interested in efficiency of the total body, and if rate

of work is not a variable factor in our experiments, we do not need to eliminate the resting metabolism from the total input.

If we are interested in efficiency of the whole body, but if rate of work *is* a variable factor, the problem becomes more complex. An argument in favor of subtracting the resting value would be that the man must go on living anyway, and that the work is not directly responsible for this cost. In the case of a real job, the "overhead" should be determined separately, and the basal metabolism for twenty-four hours would be included in a complete analysis of the input. If a man runs ten miles per hour, the basal metabolism per mile during the running is much lower than it is when he is walking at one mile per hour. If he runs ten miles per hour in making a certain trip, however, he must rest more frequently and longer. His basal requirements go on during those rests as well as during the time of running itself. If he isn't able to go as far in a day when he runs as he goes in walking at a slow pace, the basal metabolism per mile would be *greater* at the faster pace.

To carry further the interpretation of the problem of efficiency of walking, it would, therefore, be necessary to know more of the conditions under which the activity is to take place—how far the subject is to walk, how his rest is to be distributed, and so on. The strictly mechanical efficiency of walking cannot be computed from the data of Poffenberger's table because the various values were not taken from the same subjects. If we did make such a computation from the above table, we should find that the most efficient rate is slower than the four-mile-per-hour optimum indicated by the values of the last column. In fact, if the data were all from the same subjects, the optimum would be below the slowest rate shown in the table.

Metabolism in Mental Work.—Tasks like mental calculation or memorizing, which can be carried on with a minimum of muscular activity, cost the organism very little in terms of energy expenditure. The estimates vary with the conditions of the experiment, but most of them are low.[2] We shall not go into the theoretical implications of this finding here. It suffices to point out that oxygen consumption is

---

[2] F. G. Benedict and C. G. Benedict, Mental effort in relation to gaseous exchange, heart rate, and mechanics of respiration, Carnegie Institution of Washington, 1933, Publication No. 446. D. E. Rosenblum, Untersuchungen über den respiratorischen Gasstoffwechsel und Energieverbrauch by geistiger Arbeit, *Arbeitsphysiologie,* 1932-1933, *6,* 214-234. See also G. H. Rounds, H. J. P. Schubert, and A. T. Poffenberger, The effects of practice upon the metabolic cost of mental work, *Journal of General Psychology,* 1932, *7,* 65-79. The latter authors give a larger estimate than Benedict because they do not try to minimize the motor involvement of the mental tasks.

not a reliable measure of the cost of work in tasks which involve muscle activity only incidentally.

Bills [3] has argued that even though the caloric output of the body is increased only slightly by these sedentary activities, the increased energy expenditure may involve such a small area of the nervous system that it really represents intense activity at that point. He draws an analogy with an electric clock which uses little electric current, although the current is essential to its functioning. All this may be true, but the fact still remains that we cannot detect these changes in the total calorie output of the body except under very carefully controlled and unnatural conditions, and even then not very reliably. Until we know where these foci of energy release are located in the body, and have methods of measuring these local events, metabolic studies will be of little value in the study of efficient conditions of mental performance.

Barmack's [4] work supports the position we have taken here. While he finds that there is usually an increased oxygen intake at the start of a new task, what happens after that is exceedingly variable. Barmack believes that there is a tendency for metabolism to *fall* when the subject becomes bored with the work. This leads to a paradoxical situation. We should expect that the effort required to continue working would increase when the subject is bored and disinclined to work. If so, we should expect oxygen consumption to increase if it bears any relation at all to the cost of this kind of work.

Fatigue

From our common observation we would expect that fatigue would be one of the important factors to be taken into account in evaluating the cost of work. The worker on a job for which he is not fitted may grow so tired that he is willing to take another at lower pay. He does not feel that his present job pays him for the extra input which it demands. Especially in those types of work in which metabolic indices cannot be applied, it would seem that a measure of fatigue should be applied.

Our common observation is probably correct, but it is not an easy suggestion to carry out. One of the primary difficulties is that the term "fatigue" is not clearly enough defined to aid us in a search for a measure. Not only is the term vague as it appears in common usage, but it has been difficult for scientists to agree upon a common

[3] A. G. Bills, *The Psychology of Efficiency,* New York, Harper & Bros., 1943.
[4] J. E. Barmack, Boredom and other factors in the physiology of mental effort, *Archives of Psychology,* 1937, No. 218.

meaning of the term which would put their research upon a straightforward basis.

Unfortunately, *fatigue* refers to several distinct phases of the cost of work. The several items included under this heading may be loosely correlated with one another, as implied by the use of the common term "fatigue" to describe them all. At the same time there is evidence that the correlation is not close enough to permit us to measure one kind of fatigue alone, and rest secure in the belief that we have accounted for all the fatigue cost by means of a single index.

As a general statement, we may say that fatigue refers to a large number of residual effects of work. That is, it refers to any inhibitory effects of activity which carry over to the period after work ceases, and also to effects which cumulate during a period of continuous work. The only residual effects which would be excluded from the definition would be permanent changes or non-reversible changes, such as structural damage to some tissue or organ (deafness, lameness, heart disease, etc.) and those changes which represent a facilitation of performance (learning, warming-up, positive transfer).

Such a definition may be regarded as too broad to be useful. In fact, the term *fatigue* is useful only as a general designation for a broad area of research. The whole area is important, however, in any study of total human efficiency, since all the effects classed under the heading of fatigue are phases of the cost of work to the individual. It does not matter whether we decide to restrict the term to a single unitary phenomenon or not, the other factors also must be considered, whether they are considered under the heading of fatigue or under some other designation. We do not believe it is possible to restrict the term in this manner, for several reasons. For one thing, common usage is against the restriction, and we should be inviting misunderstanding. Secondly, attempts to narrow the meaning of the term have not been very successful because it has been difficult to delimit the boundaries in an unequivocal fashion.

This broad topic of fatigue covers certain important focal areas, each one of which is important, and each one of which is a very complex problem for investigation. The subdivisions of the field may be listed as follows:

1. Feelings of tiredness, weariness, or exhaustion (sometimes called "perceived fatigue").
2. A reduction in the worker's *capacity* to perform his job, as a result of previous performance of the same task.

3. Changes in other activities and capacities (including changes in physiological function and changes in ability to perform psychological activities other than the job itself).

We may think of all these phenomena as dependent upon certain fundamental bodily changes—changes in neural function, in the metabolic activities of certain organs, and so on. If it is preferred, we may consider that "true fatigue" consists of these fundamental patterns of bodily change, and that the three sets of phenomena described above are all indirect effects of the same set of ultimate factors in the body. The lack of precise correlation between various effects, such as weariness and loss of capacity, may be explained as due to the different indicators reflecting different phases of the fundamental bodily changes.

Even though much physiological research seems to reflect the above point of view that "true fatigue" is something which underlies these more superficial expressions of fatigue, it is not a point of view which is of much value for most practical studies of worker efficiency. Perhaps some day the situation will change and justify such a notion of "true fatigue." At present, however, it is only occasionally a useful concept. First of all, the items listed above as phases of fatigue are much closer to direct observability than the underlying changes in physiological function. Secondly, there is little evidence to support the point of view that there is any *single* pattern of organic change which is the characteristic basis of all the aspects of fatigue mentioned above. On the contrary, as we shall see, present evidence indicates that similar changes in capacity to perform may depend upon one set of bodily conditions under some circumstances and upon an entirely different set of conditions if the situation is changed. Even with this concept of fatigue, therefore, the term still designates a very broad class of phenomena.

Feelings of Tiredness.—Although the principal experimental studies have been concerned primarily with the capacity changes, it should not be assumed that the feeling of tiredness is unimportant. Perhaps the layman is not very likely to make such an error, but the professional writings upon the subject of fatigue sometimes leave the impression that the feelings of the worker are somehow less important than the more "objective" effects of work. Such an implication is probably not intended; it may result from the difficulty of securing a workable research approach to the problem of feelings, and a satisfactory definition of what is meant by the common-sense term "tiredness."

The reaction of the layman to this discussion is likely to be something like this: "I don't see why a definition is so important, or even necessary. I understand what others mean when they say they are 'tired out,' 'weary,' or 'exhausted.' Why not treat this as something which is self-evident? We don't have to define what we mean by 'red' or the 'taste' of an orange, so why should it be necessary here?"

Such an objection has its points, but there are good reasons why a definition is needed, even so. First of all, perhaps we *are* all referring to the same things by these terms. A scientific study of the problem requires something more, however. Suppose we are faced with the problem of which of two methods of doing a given job is more efficient. We might ask the workers to use each method and then ask them which is more fatiguing. Unless they work for a long time with each method they would not be able to assess them because the first adaptation to the method might be quite different from what happens after they are thoroughly accustomed to the conditions. Therefore we should have to require that our subjects compare degrees of fatigue from memory over a considerable period of time. Even if this were possible, it would still be impossible to say *how much* one method excels the other. If the production rate on the two methods is changing at the same time, the problem is still more complicated. We should have to ask the subjects to estimate a *quantity* of fatigue so that we could estimate the amount of fatigue *per unit* of work.

There have been some attempts to use a refinement of this procedure, in the form of a rating scale. Poffenberger, for example, presents the curves of Figure 2 derived from a study of output and feelings of tiredness.[5] The subjects rated their feelings at regular intervals during the experiment, on a scale ranging from + 3 to − 3.

Notice, however, that these results are not used as a measure of quantity, but simply as a method of comparing *relative* values at different times on the same day. And this is as far as the method could be carried.

Fatigue Regarded as Reduced Capacity for Work.—In order to avoid the above difficulties with perceived fatigue, physiologists began in the nineteenth century to perform experiments upon isolated muscles and simple muscle activities of the intact organism, using an easily observable criterion of fatigue. Their purpose was to discover

---

[5] A. T. Poffenberger, *Principles of Applied Psychology,* New York, D. Appleton-Century Co., Inc., 1942, p. 111.

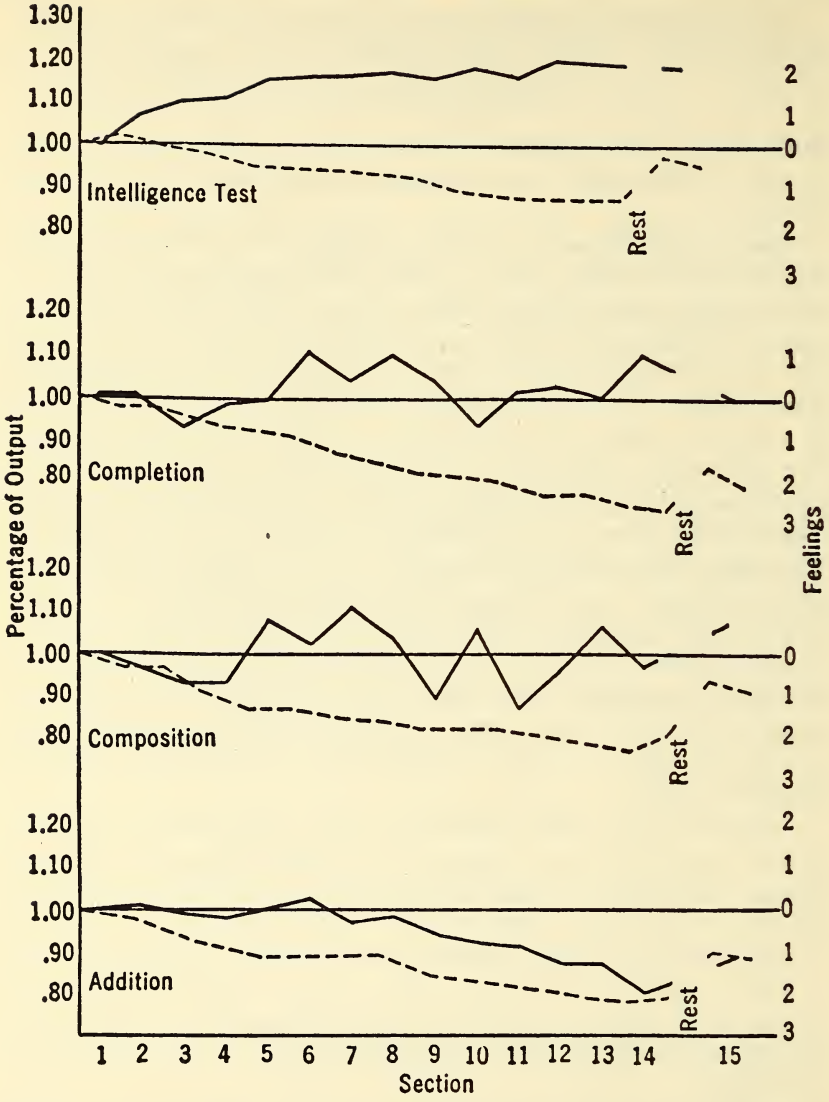


Figure 2. Relation Between Output and Feelings in Four Forms of Mental Work
Solid line represents output, and dotted line represents feeling record.
(Reprinted from Poffenberger, *op. cit.*, p. 111.)

the fundamental processes and principles of muscular fatigue, in the belief that these discoveries would lead to better understanding of fatigue in more complex activities. The criterion of fatigue was a reduced capacity of the muscle to perform, as a result of previous work.

By thinking of fatigue in these objective terms, they were able to obtain definite advances in our knowledge of muscle function and of certain aspects of muscular fatigue. We shall presently review the results of these experiments upon muscle capacity. The same concept of fatigue has been applied to many activities outside of the laboratory, activities which are more complex than those which were the subject of these physiological investigations.

As a yardstick for measuring fatigue of all kinds of work, and under all kinds of conditions, reduced capacity for work is subject to a number of difficulties. It is useful and justifiable to study fatigue in this manner when we are dealing with an isolated muscle in the laboratory, or even with carefully controlled muscular activities of an intact organism. When the same definition is applied to such complex activities as daily work in a factory or in the schoolroom, however, it is much more difficult to justify. The output may be readily observed, but it is often impossible to ascertain what has happened to the level of interest of the worker. In short, we do not know what has happened to the *capacity* of the worker, only that his *output* has changed. Muscio [6] has considered this difficulty so serious that he has proposed dropping the term "fatigue" altogether.

Another proposed solution of this problem is that fatigue be defined as reduced *efficiency* rather than as reduced capacity.[7] In other words, if the worker has to exert greater effort to accomplish a given amount of work, he is becoming fatigued. If this definition is adopted, fatigue itself would await the solution of the general problem of measuring input. Perhaps this is the only way in which these difficult problems of logic can be fully settled. At the present stage of our knowledge, however, fatigue itself refers to a group of elements of cost which have not been analyzed and separated. Until that analysis is complete, it is preferable to use the term "fatigue" as a general term for these elements, and to try to find some indicators of an indirect sort which can be used as additional measures of effort and added to the input factor in our efficiency ratio.

Under laboratory conditions, as we have said, a definition of fatigue in terms of reduced capacity to perform is often workable because capacity changes are reflected in output. With interested, cooperative subjects it is possible to use tasks which are extremely severe, and which require the subject to work nearly at capacity. It

---

[6] B. Muscio, Is a fatigue test possible?, *British Journal of Psychology,* 1921. *12,* 31-46.

[7] M. E. Bitterman, Fatigue defined as reduced efficiency, *American Journal of Psychology,* 1944, *57,* 565-572.

is also possible to keep a rather close check upon the individual subjects and their behavior. They can be questioned and expected to answer quite frankly, in so far as they are able to describe their interest and degree of effort.

Under field conditions in the factory or school, however, these conditions cannot often be met, and consequently measurement of fatigue in terms of output-decrement should be accepted only critically and with reservations. Unfortunately, there have been some who have not recognized this need in evaluating studies of industrial output.

Most of the preceding discussion was concerned with fatigue in muscular work. The precautions which are necessary in applying the concept of fatigue as reduced capacity to perform are even more important in the kinds of work which we speak of as "mental" tasks. Many muscular tasks make repeated demands upon a single set of muscles and associated nerves. It is under these conditions that we observe the most rapid and clear-cut decrease in performance. Most "mental" work is, however, extremely variable in the demands which it makes upon the body. Since these demands are less sharply localized, the pattern of performance in these tasks does not show the regular decreases which are clearly seen in more limited muscular work. In mental work, it is only when we consider an extremely artificial task in which one small operation is repeated at frequent intervals that rapid and clear-cut decrement appears. It is likely that the dissimilarities between these artificial mental tasks and normal working activities are greater than the differences between work in the ergographic experiment and other types of muscular activity.

To sum up this phase of our complex problem: Measuring fatigue by measuring the reduction in performance during work is subject to severe criticisms, but it is probably justified if used with care in the following situations:

(a) Carefully controlled laboratory situations with highly motivated subjects.
(b) Factory work or sports in which there is a heavy demand upon the muscular resources of the body.

Laboratory Studies of Muscular Fatigue

Techniques of Study.—When a muscle and its attached nerve are removed from the body of a cold-blooded animal like the frog, electrical shocks applied to the nerve or the muscle cause the muscle to

contract. It is possible to arrange apparatus which will make a graphic record of the muscle twitches, and to study the effect of a variety of factors upon the strength and duration of the contraction. These experiments have now become a common part of the laboratory exercises in college classes in physiology. It is found that the muscle which is stimulated repeatedly and frequently gradually diminishes its response until finally it will contract no longer. Fatigue is defined as the loss of strength of the muscle after these repeated stimulations.

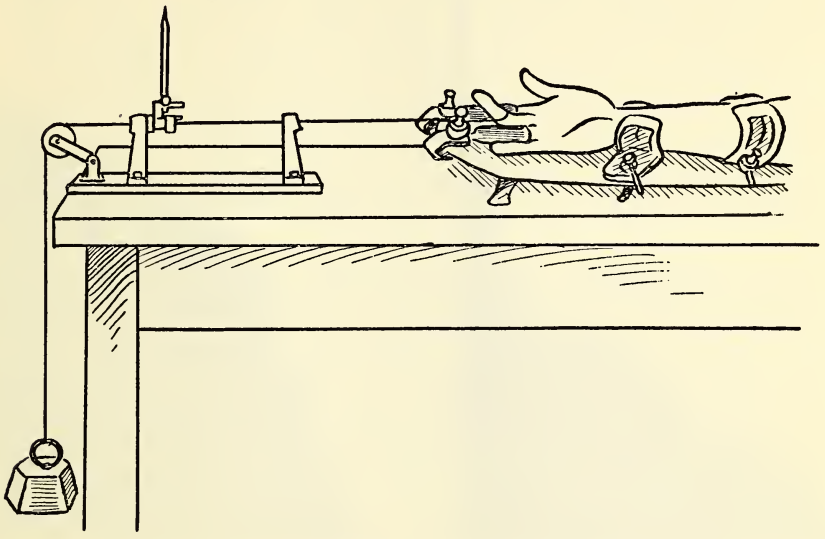


Figure 3. Finger Ergograph as Designed by Mosso

The pointer at upper left traced a record of movements upon a smoked drum (not shown). The first and third fingers are held stationary by being inserted into metal cylinders. Lifting is done by the second finger.

If the muscle is no longer able to respond at all, it is spoken of as "exhausted."

Some of the results of these experiments upon isolated muscles have played a part in developing theories of fatigue which are current among physiologists. We shall therefore have occasion to refer to these results a little later.

Similar experiments upon muscles in the intact organism were made possible by the instrument known as the *ergograph.*[8] In this instrument (shown in Figure 3) [9] a weight is attached by a cord to a

[8] A. Mosso, Les lois de la fatigue étudiées dans les muscles de l'homme, *Archives italiennes de Biologie,* 1890, *13,* 123-186.

[9] *Ibid.,* p. 130.

fingertip, the hand, or the foot. The subject is instructed to raise the weight each time a signal is given, and to continue to raise the weight as high as he can each time the signal is heard. In this way a pattern of muscle contractions can be traced upon the recording drum in a fashion similar to that described in the experiments with isolated muscle. In this case, however, the method of stimulating the muscle is the normal "voluntary" nerve impulse rather than the artificial

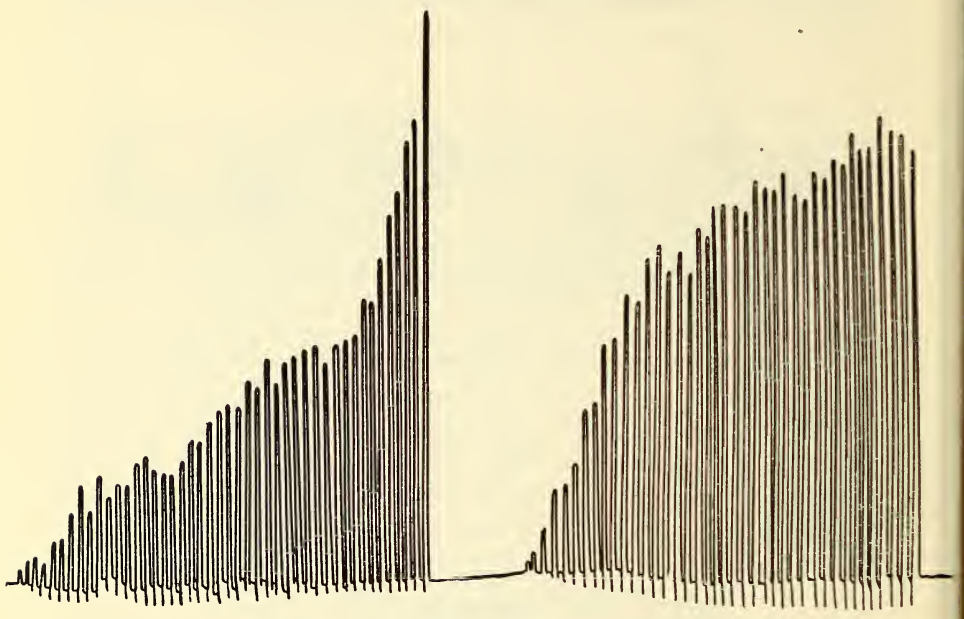


Figure 4. Records of Two Ergograms

Each stroke represents one lift and drop of the weight. The temporal order is from right to left. The two records were made by different subjects to illustrate the characteristic individual pattern found by Mosso.

electric shock. There is also another difference. In the intact body, a weight is not always lifted by exactly the same combination of muscles. There is, then, a possibility that the muscles involved in writing the ergographic record will not be the same throughout the experiment. This is partially controlled by strapping down the limb of the subject and, if the finger is to be used in lifting, restricting the movements of the other fingers.

The advantage of this approach to fatigue is obvious. There is no longer any possibility of differing meanings of the term, and fatigue becomes measurable. These advantages have been considered so important that a great deal of research upon more complicated activi-

ties, including factory work and school work, has employed the same definition in terms of decreased output.

It is important to note, however, that there are certain definite conditions under which this approach to fatigue can be made. In the case of the isolated muscle, a reduction of the force of contraction, as shown on the record, means that the system no longer has the *capacity* to respond to the stimulus. Is this also true in the ergograph? Suppose, for example, that the subject simply loses interest in the task, and reduces his output for that reason. Should this also be called fatigue? It does not seem useful to do so. Therefore we must specify that *fatigue can be defined as a loss of strength of contraction only under constant motivation.* In a great many of the ergographic experiments, this condition has been reasonably well satisfied. The subjects have been men who were interested in the problem and definitely motivated to perform with their maximum possible strength.

Evaluating the Results of Ergographic Experiments.—During the years around 1900 there was a widespread interest in research in fatigue by using the ergographic technique. Many of the results were helpful in increasing our understanding of the problem. Unfortunately, the data which are reported from the experiments are extremely difficult to evaluate because the experimenters of that day paid little attention to the possible effects of chance in determining their results, and often failed to control sources of constant errors. In recent years there has been some loss of interest in this type of research, and many of the studies which have been carried out still omit statistical analyses. This is unfortunate, since our modern methods of control in experimentation, and of statistical analysis, could clarify many things which were left obscure by these experimenters. We have, as a result, many suggestive findings, but little that is conclusive. In spite of this, many of these results have been adopted in an uncritical way by texts in physiology and applied psychology. We shall present some of these findings because of their interest, but we do so only after a preliminary warning.

Here is a typical case which illustrates this warning. The Belgian Mlle. Ioteyko examined the effect of exhaustion of the muscles of a finger of one hand upon the muscular strength of the opposite hand.[10] This experiment had a number of theoretical consequences relative to

---

[10] J. Ioteyko, L'effort nerveux et la fatigue, *Archives de Biologie,* 1899-1900, *16,* 479-535. Participation des centres nerveux dans les phénomènes de la fatigue musculaire, *L'Année de Psychologie,* 1900, *7,* 161-186.

the nature of fatigue. Out of the eighteen subjects, some showed a decrease in strength of the opposite hand, others showed an *increase,* and the remainder showed no change. Our first reaction should be to raise the question whether this result represents a simple matter of chance variations in strength of the hand, variation which would have occurred regardless of the fatigue in the right hand. Ioteyko, however, does not consider this question at all. Instead, she says that there are three types of individual. In one type, the fatigue is mostly local and in the muscles themselves, and there is no transfer of effect to the opposite side. Another type has fatigue which affects the central nervous system more severely, and therefore diminishes the strength of any reaction. The third type is "mixed." We are not concerned at the moment with these theories. There are many criticisms which could be raised against them even if we accepted Ioteyko's first assumption. The first assumption itself is not, however, demonstrated. In order to demonstrate it, it would be necessary to show that certain individuals *consistently* increase the strength of one hand after fatiguing work with the opposite hand, while others consistently decrease their strength. In other words, we need to show that we have something more than a random variation in strength unrelated to the effects of fatigue in the opposite side. This cannot be done without adequate statistical and experimental controls. The experimenters of that time were not aware of the methods which are now in routine use for answering questions of this kind.[11]

In addition to these problems of experimental method, there are certain assumptions sometimes involved in the use of the ergograph which must be clearly realized in interpreting the results. The most important is that "exhaustion" can be taken as a fixed point of reference—as representing a fixed state of the muscles, or a given "quantity" of fatigue.

Let us see how this assumption is important. Suppose that we are trying to find out what load or weight leads to most efficient performance of the finger muscles. The method would be as follows: Let the subject work to "exhaustion" a number of different times with each of several weights. We then find the amount of work done

[11] More recently, Collier has found a statistically significant crossing of fatigue in a task involving rotating the forearm at maximum speed. (R. M. Collier, The crossed effects upon voluntary movement of a unilaterally induced fatigue, *Journal of Experimental Psychology,* 1938, *23,* 26-43.) The number of subjects and the methods of analysis were not adapted, however, to determining whether there were different individual "types" of subject. The crossed fatigue resulted in decreased speed, but also in increased amplitude of movement. The greatest effect involved the speed of *reversing* the direction of movement rather than the speed of movement itself.

in each ergogram by multiplying the weight by the total distance it has been lifted. It will be found that the total amount of work varies from weight to weight, and in this manner we determine the optimal point. As we should expect, a weight which is too heavy exhausts the muscles so quickly that very few lifts will be made. On the other hand, with a very light weight, little is accomplished on each lift, and the slower fatigue is canceled by this fact.

To simplify the underlying procedure: we have assumed that the cost of work per ergogram is a constant, that a given amount of fatigue is always involved. If this is true, the conditions which produce the greatest total accomplishment in an ergogram are the most efficient conditions. This is often taken as a self-evident assumption—an axiom, but it really depends upon the question of whether exhaustion or cessation of movement always implies the same state of the organism. This must be discovered before the procedure indicated above is justified.

There is, however, evidence that the effects of completing an "ergogram" upon the organism are not always the same, and that we are not justified in saying that the cost remains constant from one ergogram to another. Of course, this evidence itself is incompletely established for reasons already mentioned, but it is enough to show that the assumption is not acceptable as an axiom which needs no testing. Here is one sort of fact which argues against the assumption: Féré,[12] one of the early workers, says that fatigue is more "profound" under some conditions than under others. For example, he finds that the *total work* done before exhaustion, increases as the rate of contraction is decreased. But he also finds that the *recovery* is more rapid when the exhaustion was reached by a series of rapid movements than it was when it resulted from a series of slower contractions.

In order to decide upon the relative efficiency of various conditions and methods of muscular work, it is therefore necessary to take into consideration a number of factors: (1) Was the motivation constant, so that exhaustion represents the same stage in ability to contract under each condition? (2) What effect does the condition have upon recovery, as well as upon the amount of work done up to the point of "exhaustion?" [13] (3) What is the effect of the condition upon a long series of ergograms separated by rest periods, as well as upon a single

---

[12] Charles Féré, Travail et plaisir, Paris, 1904.

[13] Some of the more recent experimenters have been able to produce very interesting hypotheses about the nature of muscular fatigue by analyzing variations in recovery as well as output of the ergogram itself.

ergogram or a few? The latter question concerns any cumulative effects of the work which do not show up in short term efforts; it may be extremely important.

Fatigue and Mechanical Efficiency. It might appear that measurement of fatigue and measurement of metabolic cost are simply two different approaches to the same effects of work, especially if we are concerned with work which is largely muscular in nature. At one time Atzler [14] made this assumption explicitly, stating that a method of work which permitted the greatest output before exhaustion would also be the most efficient from the point of view of metabolic cost. He cited experiments in which he had his subjects work at various rates upon the bicycle ergometer. Those speeds which permitted the greatest amount of work to be accomplished before the subject was exhausted were also the speeds at which mechanical efficiency was at a maximum during the early minutes of work. That is, the lower the oxygen consumption at the beginning of the work, the more the subject accomplished before he was exhausted.

Simonson, however, questioned the possibility of generalizing this result into an axiom.[15] In fact, he was able to show that factors other than variation in speed could produce an opposite effect. In lifting weights with the outstretched arm, the weights which were handled most efficiently from the point of view of oxygen consumption were the ones which were the least efficient from the point of view of total accomplishment in the ergogram. The results corroborated those of Atzler when the changes in efficiency were produced by changes in speed of work, but contradicted them when the load was varied.

If metabolic cost and fatigue expense are not correlated in heavy muscular work, as Simonson's experiments indicate, they must be measures of independent elements of input, and so must be measured separately. If this is true for muscular work (where there was some reasonable expectation that they would be identical), it must be even more evident when we consider tasks involving much less heavy muscular effort.

Summary of Research on Muscular Fatigue.—Because of the difficulties of research methods discussed in the preceding paragraphs,

---

[14] E. Atzler, Die Bekämpfung der Ermüdung, in E. Ludwig's *Der Mensch im Fabrikbetrieb,* Berlin, 1930, 15-37.

[15] E. Simonson, Der heutige Stand der Theorie der Ermüdung, *Ergebnisse der Physiologie,* 1935, *37,* 303-305.

views and theories of the nature and basis of muscular fatigue vary considerably. The following highly condensed analysis of the implications of these researches must, therefore, be tentative in character, although we have tried to use the more fully verified results. We shall attempt an organized summary because the only other alternative is a listing of a great number of heterogeneous results and theories which would have little meaning. Even if the following statements are tentative in character, it is hoped that they will give the reader some understanding of the scope of the problems facing specialists in muscular fatigue, and some notion of the progress which has been made to date. We must condense our summary to a very small compass in relation to the number of experiments and discussions available, because a more complete survey would require space and time out of proportion to the practical gains.

1. Loci of Muscular Fatigue. A question which has concerned most of the researchers in this field is: "When capacity to continue performing a given task has been reduced or eliminated altogether, which parts of the total mechanism are primarily responsible for the loss—the muscles themselves, the peripheral nerve connected with the muscles, or special portions of the central nervous system?" Since all these mechanisms are engaged in the performance, they are all affected. It appears, however, that the effects are more serious in some places than in others. The relative importance of the events in various parts of the total mechanism varies, however, with the conditions and type of work.

(*a*) *Local "Dynamic" Work.* We may consider first the "local" fatigue which occurs when a small group of muscles is involved, as in the finger-ergograph. It appears that the central nervous system is responsible for reduction of output before the peripheral mechanisms are seriously affected in most kinds of localized muscular effort of the *intact* organism. The muscle itself is injured only under abnormal experimental conditions, or in those experiments where the muscle is isolated from the rest of the body. This fact is of importance not only because of its contribution to our understanding of the total pattern involved in fatigue, but also because chemical theories of the basis of fatigue have in the past been based upon the isolated muscle, and are therefore concerned with an abnormal or atypical kind of fatigue, or one phase of fatigue which is of less importance than others.

After exhaustion of the power to lift a weight by voluntary contraction, the muscle will still respond if electrical shocks are applied

to the muscle directly or to its nerve.[16] It has been found, however, that after voluntary exhaustion in work with a heavy weight and a very fast pace, there is a definite effect upon the power of the muscle to contract under electrical stimulation, while exhaustion produced by lighter loads or slower contractions (slower fatigue) shows no evidence of fatigue in the muscle itself.[17] In fact, Reid found that he could not produce fatigue by direct stimulation of the muscle or its nerve unless the stimulation was repeated over a long period of time, or unless he used continuous stimulation rather than periodic shocks. This applies, of course, to the muscle in an intact organism. The rapid fatigue under this kind of stimulation in an isolated muscle, must, he believes, be due to the lack of normal blood circulation.

(*b*) *Postural Restriction—Maintained Contraction.* In static work (holding a weight continuously), as contrasted to dynamic work (repeated contractions of short duration like those referred to in the preceding discussion), the roles of central and peripheral fatigue may be different. In his comparisons of voluntary contractions with direct electrical stimulations, Reid concluded that most of the reduced capacity is a result of central factors in both static and dynamic work. Simonson, however, advanced the theory that static work with very heavy weights which produce rapid exhaustion is largely a matter of local muscular effects, while slowly developing static fatigue involves a more marked central component.[18] His evidence was largely indirect and has been questioned by more recent researchers, however. Kurbatova and Sheidin reach an opposite conclusion on the basis of the effects of rest.[19] They find that work capacity recovers much more rapidly after quick exhaustion in static work than it does in the slower development of fatigue. This recovery from rapid static fatigue takes place even if the circulation of the arm is cut off during the rest period. Massaging the muscles is more effective following slow fatigue than it is after rapid fatigue. More recovery from the rapid fatigue could be produced by some activity of the other arm than by complete rest. These conclusions would therefore be exactly the reverse of these drawn by Reid for dynamic work. In static work, the central factors are more important in rapid fatigue than they are in slow fatigue; in dynamic work the central factors are most significant in *slow* dynamic work.

---

[16] Mosso, *op. cit.*

[17] C. Reid, The mechanism of voluntary muscular fatigue, *Quarterly Journal of Experimental Physiology,* 1928, *19,* 17-42.

[18] Simonson, *op. cit.,* p. 330.

[19] I. N. Kurbatova and Y. A. Sheidin, On the question of fatigue in static work, *Uchenie zapiskie,* Leningrad University Publications, 1938, Vol. 6, No. 23, 149-163.

Little can be said at this time about the nature of the changes in the central nervous system that are responsible for diminishing muscular output; the role of the central nervous system in fatigue was discovered by a process of elimination. The portions of the nervous system involved might be sensory centers, motor centers, or coordinating connections. There are, of course, various theories, but the evidence is fragmentary, and this is a problem which needs much further analysis. It is known that fatigue of the nerve fibers themselves is a negligible factor under normal conditions. Solution of the problem of the intimate nature of the changes involved awaits better understanding of the physiology of the nervous system in general.

At any rate, the multiplicity of factors involved even in what appears to be a very simple kind of "local" muscular fatigue, and the variability of the factors with changes in conditions, should make it clear that practical recommendations upon the control of fatigue cannot be worked out by derivation from any general theory of the nature of fatigue. Even if the general theory were well established, we should still need to know what combination of factors is important in any particular case before fatigue could be controlled.

In more generalized muscular work of a heavy kind (dromal tasks), such as running, bicycling, carrying heavy loads, and the like, the nature of the basis of fatigue is probably still more complex. It has been concluded that in rapid exhaustion in heavy work, the primary factor is a failure of the circulatory and respiratory mechanisms. The central nervous system may be affected by the deficiency of oxygen, however, and consequently may again be a central factor in the cessation of effort.[20] In general muscular exertions which are of the slower kind—muscular drudgery in which a steady effort is continued for longer periods of time, as in hiking, digging ditches, other common labor, and the like—the mechanisms are probably different from those involved either in "local" fatigue or in rapid exhaustion in heavy work. It is, in fact, probable that the mechanisms are so distinct that the basis of fatigue could not even be suggested from experiments upon the other kinds of work.

2. Rest and Recovery. Some of the factors which affect the rate of recovery have already been mentioned in discussing the loci of fatigue. Other facts, of interest in their own right, will be summarized briefly at this point. Amount of recovery is usually measured in terms of the total amount of work done when the work is taken up again following a period of rest. That is, following a given

[20] Simonson, *op. cit.*, p. 340.

period and condition of rest, the task is resumed, and the subject works until he reaches complete exhaustion. In the case of dynamic work, the amount of work done during this second period of work is obtained by multiplying the weight by the total distance through which it is lifted—the sum of the heights of all the contractions.

Using this method, it was found that the later portions of a period of work are much more costly than the initial period. Maggiora, for example, measured the time required for complete recovery of the activity, *i.e.*, the time at which the second ergogram showed the same total output as the first.[21] In another series of experiments, he allowed the subjects to work only half as long as was normally necessary to reach exhaustion, allowed them to rest, and then determined the length of rest necessary to produce a normal ergogram. Under these latter conditions the time required for complete recovery was only one-fourth that required for complete recovery following exhaustion.

The recovery following exhaustion does not follow a smooth trend, but appears to be cyclical. A short rest (ten to fifteen minutes) may sometimes produce more recovery, as measured by output, than a longer period.[22] Whether or not this stimulation of performance following a brief rest is to be considered true recovery, or some kind of spurious effect, depends upon our definition of "recovery." The definition used in these experiments was stated above. In terms of this definition, the subject does recover, and then goes into a slump again before attaining permanent return to the normal condition. It is still possible, however, that the mechanism involved in this first rapid recovery is different from that which produces the final and slower return to normal capacity. It is also possible that performance is more costly after the brief rest period, even though this cost is not reflected in performance at the time.

Recently, Sharp has thrown some interesting new light upon the rapid and temporary recovery in the first few minutes after ergographic work.[23] By analysis of muscle potentials from the active muscles during the rest period, he found a spontaneous recovery of muscle tension at about the time of the first temporary recovery in work capacity. The pattern of potentials is shown in Figure 5, taken from Sharp. It would appear that muscle tension, or *tonus,* is cor-

---

[21] A. Maggiora, Les lois de la fatigue étudiées dans les muscles de l'homme: II, *Archives italiennes de Biologie,* 1890. *13,* 187-241.

[22] Charles Féré, Travail et plaisir, Paris, 1904. C. W. Manzer, An experimental investigation of rest pauses, *Archives of Psychology,* 1927, No. 90, 1-84.

[23] L. H. Sharp, Effects of residual tension on output and energy expenditure in muscular work, *Journal of Experimental Psychology,* 1941, *29,* 1-22.

related with the capacity of the muscles to perform. What this tension means with respect to the state of the mechanism of action, and what brings about the change in tensions, still remain to be discovered, but Sharp's findings represent one more clue to the nature of the process of recovery.

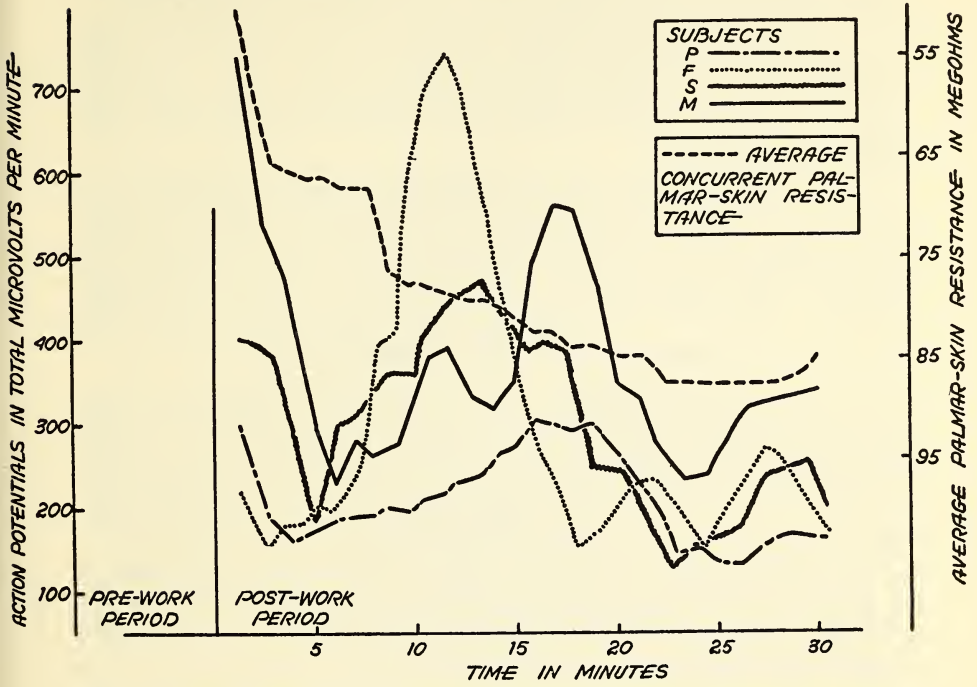


Figure 5. Skin Resistance and Muscle Potentials Following Ergographic Work (From Sharp, *op. cit.*, p. 7.)

Reid believes that the process of recovery is influenced by proprioceptive (muscle sense) stimuli in the muscles, even when the principal locus of fatigue is in the central nervous system.[24] For example, he finds that electrical stimulation of the muscle during the "rest" period —*i.e.*, when the muscle is not being used in voluntary contractions— slows down recovery. His interpretation is that the sensory impulses maintain the central fatigue condition. In other words, the stimulation of the muscle was such that it would not produce appreciable fatigue of the muscle itself, and therefore should not affect recovery on that basis. It may well be that Sharp's findings on tension, and Reid's conclusions, are both related to a similar factor in recovery.

[24] Reid, *op. cit.*

3. Effect of Load and Rate of Performance. For a constant rate of work—a fixed number of lifts of the weight per minute —there is an optimal load. Weights heavier or lighter than this optimal load will not permit the subject to accomplish as much before he reaches exhaustion. With the lighter weights, fatigue progresses more slowly, but not so much work (as measured in foot-pounds) is accomplished in each lift of the weight. For loads which are too heavy, fatigue is so much more rapid that it more than overcomes the advantage due to the greater accomplishment in each lift of the weight. In general, the slower the rate, the heavier the load for the optimum combination.

With a fixed load, the total accomplishment before exhaustion, increases as the rate is slowed down. It is possible to leave sufficient time between lifts so that the work can be continued almost indefinitely.[25] Ultimately, however, there will come a point where the subject is incapable of continuing. Performance may remain constant for a long time and then suddenly drop to zero.[26]

4. Fatigue and Spread of Activity. As fatigue increases, there is a spread of activity to other muscles of the hand. Ash found, for example, that it was possible to record contractions in the fingers which were not actively lifting the weight. As the ergogram progressed, contractions of these muscles became greater and greater.[27] If the middle finger is "exhausted" while the other fingers are strapped in place, the weight can again be lifted if these unused fingers are released. Ash concludes from this that fatigue involves a loss of fine control over these muscle co-ordinations.

5. Theories of the Chemical Basis of Fatigue. While most physiological texts and general summaries of the subject of muscular fatigue customarily devote a great deal of space to the chemistry of muscle contraction, the importance of this topic for our purposes is quite small. We shall seek only to place this topic in relation to the other phases of fatigue already discussed.

A number of facts derived from the study of fatigue in the isolated muscle, and some of the findings related to fatigue in the intact organism, led to the development of theories of fatigue in terms of "fatigue substances." These were end-products of the process of muscular contraction which were held responsible for the loss of

---

[25] Maggiora, *op. cit.*

[26] Féré, *op. cit.*

[27] I. E. Ash, Fatigue and its effect upon control, *Archives of Psychology*, No. 31, 1914.

capacity of the muscle to perform. Lactic acid is one of the substances which has received considerable attention from this point of view.[28] Increased acidity was supposed to affect the nerve junction with the muscle so that it no longer transmitted impulses; in higher concentrations, to affect the power of the muscle itself.

There is no doubt that an understanding of the chemistry of muscular contraction has its place in completing the total picture of the economy of activity. In heavy muscular work, the extra oxygen taken into the body is employed in oxidizing these chemical products of activity and in restoring the muscle to its normal chemical state. In the isolated muscle, and perhaps in extreme conditions of the intact organism, these activity products are probably in sufficient concentration to decrease the capacity of the muscular apparatus to perform. Even under more normal conditions, products such as lactic acid spread to other parts of the body through the blood stream, and may have some indirect effects in this manner.

The production of lactic acid is now considered, however, only one among a great many different bodily events involved in muscular fatigue. Even in isolated muscles it is probably not the sole factor leading to a reduction in capacity. As we have seen, the emphasis has shifted from peripheral muscular events to the central mechanisms of control in the nervous system.

Summary of Muscular Fatigue.—There is no doubt that research into muscular fatigue has made progress. A part of the picture has begun to emerge, although its details are still hazy. It is a gain to know that muscular fatigue, even under the simplest conditions, cannot be subsumed under some simple theory or formula. The definition of fatigue in terms of reduced capacity to perform, while it is difficult to apply in a factory or under other practical circumstances, is workable and useful for physiological investigations into the basis of fatigue.

In the light of the results we have mentioned, it cannot be assumed that fatigue in mental work will be understood by merely extending the conceptions derived from muscular work. Even in muscular work there are probably many different conditions which would be classed as "fatigue states" in that they reduce capacity for performance, and in that these states result from previous exercise. How much these states have in common beyond these gross features still remains to be determined. Similarly, it is likely that still dif-

---

[28] See Atzler, *op. cit.* in *Ergebnisse der Physiologie,* 1938, *40,* 1939, *41,* and Simonson, *op cit.,* for general reviews of this topic.

ferent states underlie fatigue in mental work, and we know still less, if that is possible, about what they have in common with the conditions underlying muscular fatigue.

It seems likely that there are several important bodily changes which occur in a variety of combinations in various fatigue conditions. We cannot assume in advance, however, that fatigue in mental work is nothing more than a new combination of the same factors involved in muscular fatigue. It is much safer to approach problems of mental fatigue directly, without preconceptions. If some of the factors involved in mental fatigue are similar to those in muscular fatigue, the fact should appear in the results of research. If there is little similarity, the direct approach to mental fatigue will have advanced us further, since it would prevent our embarking upon a dead-end trail.

As for measuring fatigue as an element in the cost of muscular work, decrement of performance in highly motivated workers is still the only useful indicator. This, as we have pointed out, has its drawbacks when applied to the practical setting because of variation in motivation. Nevertheless, research into the physiological basis of fatigue has not provided us as yet with any adequate bodily test for determining the degree of fatigue that results from a given task. Such a test may ultimately be perfected, but meanwhile we must make the best of output decrement in the study of muscular work.

Chapter 4

MEASURING THE COST OF WORK: II. FATIGUE IN SEDENTARY WORK

Kinds of Sedentary Work.—As mentioned previously, it is conventional to divide all work into the two main classes of "mental" and "muscular." We have found that it is still useful to adopt a two-fold classification for purposes of chapter division, even though the traditional division of the two classes is not very satisfactory. In the preceding chapter we adopted a revised conception of the class of muscular work, including only those classes of work in which the central feature of the task is the development of muscular force. In other words, it is not merely that muscular activity is involved in the work, but that the stress is upon the human body as a machine for performing work in the physical sense ("dynamic" work) or to maintain equilibrium of a physical system ("static" work).

We have indicated that the measures of cost of work which are useful for studies of muscular work are of doubtful value when applied to other tasks commonly described as "mental." Before taking up the measures of cost of work for this latter kind of task, it will be well to consider the kinds of activity that are to be studied. So far we have characterized *sedentary* work primarily in a negative way—the class of work which is left after muscular work has been abstracted. No term is entirely adequate to describe this class, but the term *sedentary* seems preferable to *mental* because many of the tasks we shall include are not commonly thought of as "mental," and the implication that *any* activities take place "in the mind" is certainly to be avoided. All activities are bodily activities.

We offer here a tentative classification of types of sedentary work. The classification is given here mainly to indicate the *kind* of subdivision which may be necessary in order to bring greater meaning into our facts. This particular classification may not be the most useful one for other purposes. It is especially likely that the headings are too broad and that they need further refinement. Nevertheless we believe that it is helpful to bear such a pattern in mind as we examine the results of researches which investigate the cost of work in the general area of sedentary or mental work.

TABLE 7

KINDS OF SEDENTARY (MENTAL) WORK *

1. Problem Solving (With minimum muscular and sensory involvement.)
   e.g., calculation, solution of mathematical problems, composing, planning, directing the work of others, supervising, socialized tasks such as selling.
2. Continued Sensory Adjustment (Primarily visual tasks.)
   e.g., proofreading, visual inspecting, radio code reception, piano tuning, and the like. Reading—under difficult conditions or when long continued.
3. Motor Skill (With patterning and accuracy of movement as the central core, and force only an accessory feature.)
   e.g., typing, drawing, various machine operations, assembling machines, woodworking, sewing, acting, speech-making.
4. Sedentary-muscular (Light muscular tasks with little skill or control involved.)
   e.g., watchman, crossing guard, machine feeder, or any task where the main requirement is that the proper movements be made at the proper time. The movements themselves have few elements of skill or of force involved.

* Modified from T. A. Ryan, Varieties of fatigue, *American Journal of Psychology*, 1944, 57, 565-569.

The distinctions between these classes of work should be evident from the examples given. Many tasks may consist of alternations between the different types (including muscular), or of various combinations of the above features. For example, in long-continued typing, postural factors and even localized muscular strain of the fingers may enter. These features alone, however, would be considered only incidental elements of the task in most cases, and the task would be classified according to the main features.

We shall see that laboratory studies of fatigue and effort have stressed Types 1 and 2—problem solving and continued sensory adjustment. Occasionally a simple kind of motor skill has been investigated, not as a separate problem but grouped with Type 1 as an instance of "mental" work. Thus tasks involving motor skill have not received the attention which is consonant with their importance in practical jobs, except for gross statistical analyses under factory conditions. The situation for this kind of work is therefore very similar to that for the "steady grind" mentioned in the chapter on muscular work.

Investigations of continued sensory adjustment have been restricted to the problem of visual fatigue. Here the investigation has

been along special lines, and in consequence there is little confusion with other kinds of work. In fact, the problems and methods are so specialized that we shall not discuss them in the present chapter, reserving this topic for a later analysis of the problems of lighting (Chapter 5).

When all the facts are in, it may be found necessary to regard several of these classes as distinct and independent. Work involving muscular skill, for example, may be found to affect the organism in a manner quite distinct from problem solving, even though both lay principal stress upon control, direction, and timing of activity. Perhaps too, the steady grind mentioned as a kind of muscular work belongs with these skilled tasks rather than with the other forms of muscular work. Another possibility is that some tasks involving muscular skill resemble problem solving, others the steady grind. The various kinds of work called sedentary are therefore grouped together only tentatively, pending further information and following the groupings inherent in the thinking of past researchers.

In spite of these doubts of the basic scientific validity of the two-category scheme, the class of sedentary tasks does possess two attributes which are worthy of note. First, all are tasks in which the energy cost of work is an inadequate indicator of the total cost of work to the worker. Second, the class includes the vast majority of tasks which are important in the modern working world. As time goes on, muscular tasks assume smaller and smaller importance in industry, maintaining their position only in the world of sports.

Fatigue and Decrement in Sedentary Tasks

In the preceding chapter we discussed some of the difficulties inherent in the use of output decrement as a measure of fatigue. These difficulties are, if anything, even more troublesome in studies of sedentary work. Nevertheless, decrement has been used in studies of fatigue in sedentary work of many kinds both in the laboratory and in the practical setting of the factory or school. Our first task will therefore be a more detailed analysis of the value of output decrement as an indicator of the cost of sedentary or mental activities. After surveying the results of laboratory studies of decrement, we shall look at some of the analyses of work curves in factory work.

Factors in Decrement of Sedentary Work.—It was hoped that the investigation of the ergographic task under controlled conditions would lead to general principles of muscular fatigue which could

then be applied to more complex activities in daily life. Similarly, several experimenters have studied tasks of the kind roughly classed as "mental" in a hope of finding rules which could be applied to the office, schoolroom, or factory.

The method of experimentation is essentially similar in the two cases. In the studies now to be considered, the subjects are asked to cancel letters in a page of text according to specified instructions, to add long columns of numbers, to react quickly to signals, and the like. With these tasks it is found that the rate of decrease in performance is much slower than it is with the ergograph. In fact, after several hours on some of these tasks, highly motivated subjects may have slowed down only slightly. For this reason it is much more difficult to find clear-cut changes which accompany the progress of fatigue.

Perhaps the greater difficulty in following the effects of prolonged mental work is due to the fact that there is a strong possibility of confusing several distinct phenomena. First, there are probably two or more different kinds of fatigue which change their relative importance with changes in working conditions. It is likely, for example, that a strong emotional strain going along with the work is one thing responsible for the feelings of tiredness of the worker. The bodily effects of this factor are probably not the same as the effects of the work itself, unaccompanied by their emotional factor. This factor is given more extended consideration in Chapter 9. Then, too, there is a muscular fatigue which enters in because of the restrictions upon movement which are required by the task.

In addition to the three factors already mentioned, factors which are commonly grouped under the heading of fatigue, there is a fourth element which has been confused with them in certain experimental studies of work decrement in mental tasks. Yet it is considered by the worker himself as quite distinct from fatigue. This is the factor of monotony or boredom. We shall discuss monotony later in the general context of motivation in work, rather than in the current section.

A fifth important factor affecting the pattern of output is practice or learning, which tends to counterbalance the effect of fatigue. In the early stages of an experiment, practice can have such a marked effect that performance improves continuously throughout the course of a day, in spite of any fatigue which may occur. One method of controlling this factor in studies of fatigue is to start the main part of the experiment only after the subject no longer shows evidence of improvement due to practice. There is often a long wait before this

stage is reached, and even so, it is always difficult to be certain that the practice effects have been fully eliminated.

Various tricks in the analysis of data have been designed in order to allow for the practice effect. One device is to compare the low point of performance in a given session with the maximum point of the succeeding period of work, following an extended rest.[1] The assumption would be that the rest is long enough to eliminate the effects of fatigue, while the effects of practice carry over to the fol-

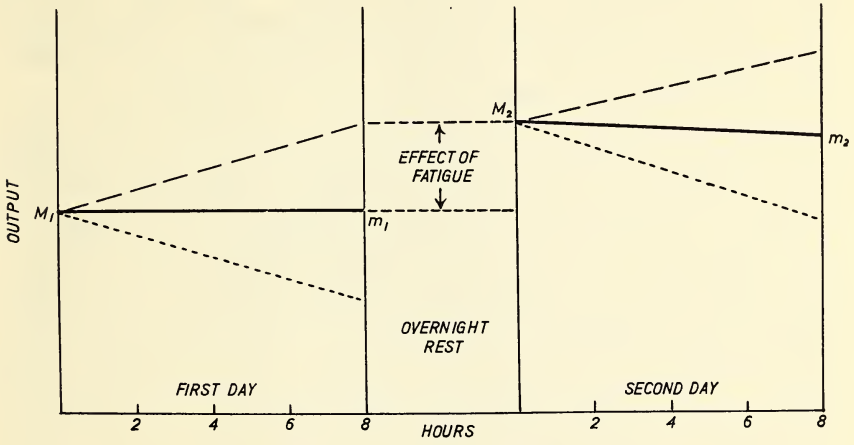


Figure 6. Hypothetical Effects of Fatigue and Practice

When practice covers fatigue effects, fatigue may be measured by (M_2-m_2) instead of actual decrement (M_1-m_1). For simplicity, this drawing represents the case where there is no warming-up effect.

lowing session. This method evidently does not allow for possible loss of skill or forgetting during the rest interval.[2] Thus the decrement in performance, computed in this manner, would be less than the actual decrement in capacity occurring during a given session of work. (See Figure 6.)

Other methods are still more complex and involve a great deal of experimentation devoted to the securing of control over factors like practice. Since most of the methods involve some assumption con-

[1] E. L. Thorndike, Mental fatigue, *Journal of Educational Psychology*, 1911, *2*, 61-80.

[2] T. Arai, Mental fatigue, *Teachers College Contributions to Education*, 1912, No. 54, p. 85.

cerning the nature of forgetting during a rest interval, it would appear much better to work with highly practiced subjects whenever that can be done. Highly practiced subjects would exhibit greater uniformity of performance and more reliable results than beginners, as well as provide for more adequate control of practice effects.

There are also other factors which tend to counteract the influence of fatigue upon decrement. Examples of these are the "warming-up" at the beginning of work, and "spurts," especially when the end of the work period is anticipated by the subject. These factors can be allowed for with relative ease, although they must be recognized and taken into account in any study of fatigue.

Because of all the factors mentioned, there are few results among these studies of mental work which are clear enough to be of value in understanding the problem. Decrement in performance can be produced experimentally, although frequently the decrement is very small relative to the feelings of fatigue reported by the subjects.[3] In a continuous alternation of tasks (tests of reaction, co-ordination, tapping, color naming, etc.) Hollingworth found some test performances improved during a twelve-hour day, while others showed decrease.[4] Even the decreases were relatively slight in amount (17 per cent loss in opposites was the greatest decrement). Practice was controlled, but there was no continuous effort in any one task. Our earlier discussion of feelings of fatigue referred to Poffenberger's curves of performance and feelings for various tasks. Here some performances *improved* during the five and a half hours of the tests, an improvement attributed by Poffenberger to practice.[5]

These examples serve to show the great variability of changes in output with varying conditions and tasks even under laboratory conditions. Robinson [6] has attempted to state the conditions which lead to decrement in a mental task, and in an experiment [7] which he carried out with Bills has demonstrated the importance of two of these factors in producing a rapid decline of performance. These two factors are *homogeneity* and *conflict.*

---

[3] See Arai, *op. cit.,* for examples of decrement in long periods of arithmetical calculation.

[4] H. L. Hollingworth, Variations in efficiency during the working day, *Psychological Review,* 1914, *21,* 473-491.

[5] A. T. Poffenberger, *Principles of Applied Psychology,* New York, D. Appleton-Century Co., Inc., 1942, pp. 110-111.

[6] E. S. Robinson, Principles of the work decrement, *Psychological Review,* 1926, *33,* 123-134.

[7] E. S. Robinson and A. G. Bills, Two factors in the work decrement, *Journal of Experimental Psychology,* 1926, *9,* 415-443.

The first of these factors is evident in ergographic work as well. In fact it is commonly said that the reason that local muscular exercise shows rapid decrement, while many mental tasks show little loss, is due to the homogeneity of the ergographic task. Almost the same movement is carried out repeatedly with only brief time intervals between the reactions. Essentially the same peripheral organs are involved, and it may be that the central neural processes undergo little change from one reaction to the next. Robinson and Bills found that a "mental" task which made similar demands would also produce a very rapid decrement. For example, the repeated writing of the same letter or pair of letters leads to a rapid reduction in output.

The decrement in this co-ordinated task is different, however, from the decrement in ergographic work. Instead of a gradual diminution of *strength* of movement, performance is *slowed down* in writing as the result of frequent "blocks" when the subject is unable to respond at all for a brief time. Between blockings there seems to be little reduction in the speed of movement. The movements themselves are rapid, but they are interrupted. The effect of homogeneity upon this co-ordinated skill is quite different, therefore, from that in ergographic work, and only in a very general way can it be said that homogeneity is a common factor in the two types of decrement. Another way in which the effect of repetition differs in the two kinds of task is that the letter-writing declines rapidly for a time, and then tends to level off. Other activities which involve speed as the measure of performance have also shown a similar pattern. It is clear that such tasks are not at all comparable to a task where performance is measured in terms of strength, while speed is controlled by the experimenter. Therefore, while it can be said that homogeneity does lead to decrement in a mental task, it is impossible to draw a parallel from its effect upon muscular work.

Bills has also cited other results of experiments upon mental work as illustrations of the role of the factor of homogeneity.[8] He mentions Poffenberger's finding that several hours of adding pairs of digits led to a 20 per cent loss by the end of the period, while answering questions on intelligence tests for the same time showed an *improvement* of 20 per cent. The greater variability of activity in the intelligence test is presumed to account for this difference.

Homogeneity, then, is quite obviously one of the characteristics of a task which leads to rapid decrement. It is possible, however, as our comparison with ergographic work indicated, that the effect of ho-

[8] A. G. Bills, *The Psychology of Efficiency,* New York, Harper & Bros., 1943.

mogeneity may vary with the conditions and with the tasks involved. Although it was apparently not a problem in the experiments we have been discussing, homogeneity may also have an indirect effect by producing monotony and reduced motivation, as well as the more direct effect of reducing the capacity of the organism to go on with the task. Bills, for example, discusses the effects of homogeneity upon blocking in terms of a "refractory phase" of the mechanism, a term which implies a relatively simple neural effect of the task. He does not, however, make a distinction between boredom and fatigue, limiting himself to the purely "objective" concept of work-decrement.

The other factor tested by Robinson and Bills, *conflict,* is largely eliminated in the simple kinds of muscular action involved in experiments upon fatigue. Conflict refers to the effect of competition between different possible responses at various stages of a task. Bills gives the example of a subject who is asked to add and subtract alternately. Here the decrement is more rapid than it would be in continuous addition, although it is possible to become accustomed to such a procedure.

While it may be useful for some purposes to state that these are general principles of work-decrement, at the same time it seems that this knowledge is of questionable value in increasing our understanding of the problems of fatigue. Let us note that Robinson and Bills reported that the effect of homogeneity was reduced if the subject became "absent-minded." This is a parallel to the fact that habituation reduces the influence of conflict. It may therefore mean that the rapid decrement is due to the fact that the subject is required to do something which is new and unnatural. Perhaps, then, both factors could be reduced to the well-known principle that new sequences of activity or new co-ordinations of movement are more fatiguing than activities which have become more habitual in their course. This argument could certainly be applied to the types of task which Robinson and Bills employed in their experiments to demonstrate the roles of homogeneity and conflict. Whether it would also apply to other cases, such as the effect upon addition which Bills cites from Poffenberger, might be argued. Perhaps in these other cases, however, there is a greater possibility of boredom's entering the picture.

Output Decrement in Factory Work.—We see that decrement in performance with time does occur in laboratory studies of sedentary tasks, although it is variable and less predictable than the decrement in muscular tasks, and we are doubtful about how clearly it reflects diminishing capacity. Before we can evaluate output decrement as

a useful index of fatigue, we must also consider the course of output through the working day in the normal working situation.

Some of the factors which reduce the adequacy of the output indicator become more troublesome in studies of factory work. The motivational pattern is much more complex, especially if the worker is paid upon a piece-rate or bonus basis. Practice, however, should be less disturbing, since most workers would be at a stage where little improvement could be expected.

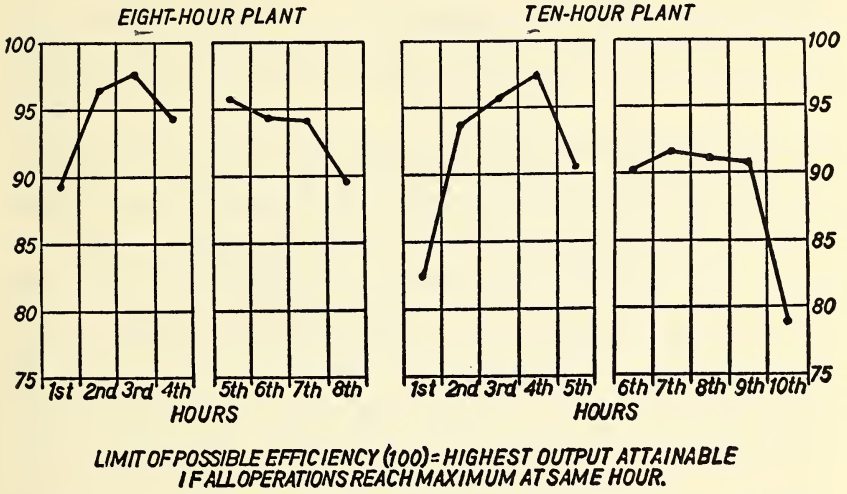


Figure 7. Composite Hourly Performance Curves for Two Metalworking Plants (From Goldmark, *et al.*, *op. cit.*, p. 74.)

Writers upon the subject of fatigue sometimes present a curve of hourly production which is considered to be "typical" of output in factory production. One basis for considering such a pattern as typical appears to be the average output curve reported in a study by the U. S. Public Health Service.[9] This curve, shown at left in Figure 7, is a composite derived from a large number of observations upon different workers and different kinds of factory work—all within the same automobile plant. This curve, however, results from canceling out various tendencies which appear in the curves of specific occupations or more homogeneous groups of occupations. In Figure 8 we have reproduced several of these more specific curves from

[9] J. Goldmark, M. D. Hopkins, P. S. Florence, and F. S. Lee, Studies in industrial physiology: Fatigue in relation to working capacity: 1. Comparison of an eight-hour plant and a ten-hour plant, U. S. Public Health Service, Public Health Bulletin No. 106, 1920.

which the composite is derived. The "typical" pattern of declining performance toward the end of the day appears most clearly in the heavier types of muscular work. In lighter tasks, it is by no means so clear, and sometimes it does not appear at all.

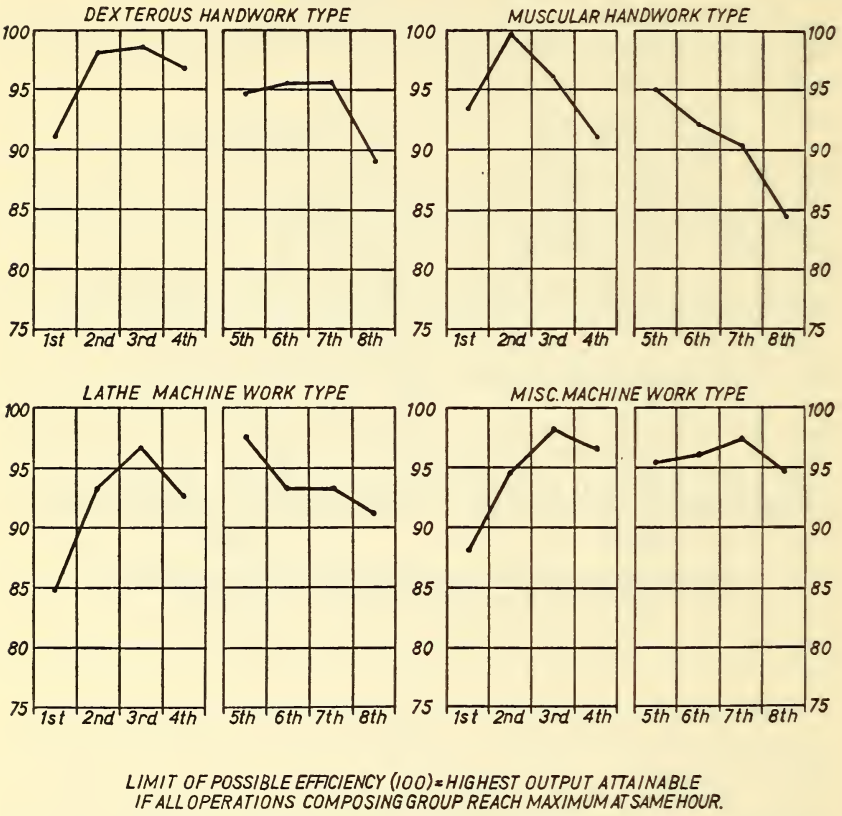


Figure 8. Composite Hourly Performance for Different Classes of Work (Eight-hour Metalworking Plant)

(From Goldmark, *et al.*, *op. cit.*, pp. 33, 45, 57, 67.)

As the writers of the report point out, there are a number of factors which tend to mask any possible influence of declining capacity for performance. Among others, they mention the factor of stereotyping. That is, the worker sets himself a constant daily quota. As an example, these authors mention that each of sixteen different workers drilled exactly 3600 pieces on every night of a given week. Variations in hourly rate are frequently a function of the worker's goal to maintain the daily average, rather than a function of fatigue.

Another finding of the same research group is that rhythmic operations tend to be more constant in rate than do more irregular tasks.

Another instance of the lack of correspondence between output and fatigue can be found in the curves of the same study of factory output. Figure 7 shows the composite curve already referred to for an eight-hour plant and another for a ten-hour plant, for essentially the same classes of work. If fatigue is responsible for decreased performance, and if fatigue is cumulative, it is difficult to see why the final decline occurs in the eighth hour in one case and does not appear until the tenth hour in the other. In miscellaneous machine work, in the ten-hour plant, the high point of production for the afternoon came in the next to the last hour, that is, the ninth hour of the day.

In practice, output curves may be of any shape; many factors enter into their determination. In several instances even simple factory tasks showed improvement throughout the working days.[10] This is interpreted as "practice" improvement, although the data were not analyzed on a day-to-day basis to see whether the workers were actually improving in the long run, or whether the improvement continued to take place every day, with a loss over night. Practice is not the only possible interpretation of progressive improvements during the day, however. Link has interpreted the increase in output in inspecting small shells (Figure 9) as the result of the monetary incentives provided.[11]

Other investigators have found curves with a low point somewhere in the middle of the work-spell (morning and afternoon). This is interpreted as the result of severe monotony in a simple repetitive task.[12] We shall question the generality of this association in our later discussion of the topic of boredom. Nevertheless, here is another exception to the "typical" curve. It would be difficult to argue that no fatigue is involved, simply because other factors have been so effective in producing an "abnormal" curve. If such reasoning were allowed, it would be possible to eliminate most "fatigue" by simply paying an extra bonus for keeping the production up during the later hours of the work-spell.

Rothe has recently collected detailed data upon the output curves of butter wrappers, giving us another example of the difficulty of in-

---

[10] H. M. Vernon, *Industrial Fatigue and Efficiency,* New York, E. P. Dutton & Co., 1921, p. 14.

[11] H. C. Link, A practical study of industrial fatigue, *Journal of Industrial Hygiene,* 1919, *1,* 233-237.

[12] S. Wyatt, J. A. Fraser, and F. G. L. Stock, The effects of monotony in work, Industrial Fatigue Research Board (Great Britain), 1929, No. 56.

terpreting these curves in terms of fatigue.[13] Composite curves for groups of workers, and composite curves for the same worker over a period of time, were compiled. It was found that these individual curves may take any one of a number of different forms. The working pattern of the same individual fluctuated widely from day to day, so that there was little predictability of the pattern for a given indi-

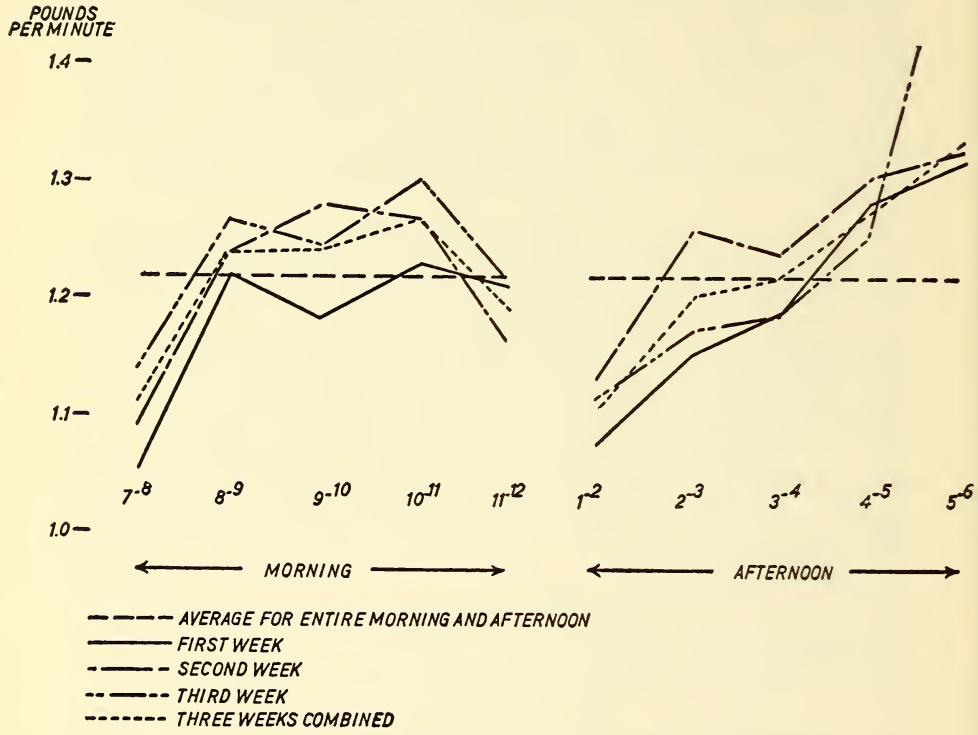


Figure 9. Hourly Performance Curves for Visual Inspection of Cartridge Shells
(From Link, *op. cit.*, p. 236.)

vidual. Some relationships did appear, however. There was a tendency for the patterns of different individuals to be similar for a given day. Composite curves for individual workers, but covering a period of several days, were interrelated. That is, two workers might show similar composite curves, even though an individual worker was not consistent with himself from day to day. Although these composite curves showed a low period of production toward the end of the afternoon, it is not stated how much production is reduced during this

[13] H. F. Rothe, Output rates among butter wrappers: I. Work curves and their stability, *Journal of Applied Psychology*, 1946, *30*, 199-211.

time. If decrement does occur in this occupation, it is evident that it can be discovered only by studying average results from a large amount of data.

Another possible index of the effects of fatigue would be a comparison of the average hourly rate in the morning and afternoon. Vernon has collected extensive data upon this relationship. In his data, production is only 2 per cent lower in the afternoon, even on those jobs where the working day was ten hours long.

These examples are enough to indicate the state of affairs. Perhaps many jobs do show a bona-fide fatigue decrement toward the end of the day (a decrement which is not due to decreased motivation or to the worker having completed a quota which he has set for himself). Even if we grant that possibility, there are sufficient exceptions to the "typical" pattern to make it a doubtful index of fatigue. If the typical pattern appears, we still have difficulty in showing that the decrement is not a result of lowered motivation. If the typical pattern fails to appear, it cannot be taken as evidence of lack of fatigue. We can only say that the output criterion has failed as an indicator of variations in working capacity.

Possible Approaches to a Study of Fatigue in Complex Tasks

It is clear from the laboratory and factory studies of work decrement in complex tasks or "mental" performance, that fatigue can be studied in this manner only for carefully designed tasks which may be quite unlike normal working activities. Even here the amount of decrement may be slight and difficult to measure reliably. There are at least four possible ways out of this difficulty:

1. Limit experimental studies of fatigue to the laboratory, using only tasks which are known to show reliable decrement. Thus, if we wished to study the effect of nutrition, say, upon fatigue, we should measure its effect only upon those tasks which may be found to exhibit the property of rapid decrement (*e.g.*, continuous addition, naming colors as rapidly as possible, canceling letters from a page of pied text).

The obvious difficulty with this approach is that it assumes that all fatigue is alike, or at least that all "mental" fatigue is alike. We should have to assume that a similar effect of a change of diet would be found for fatigability in other tasks, even though it cannot be measured. It overlooks the possibility, indeed almost the certainty, that the effects of work differ with the task. Consequently we should

not expect that a change in vitamin content would have the same effect upon the executive who is tired after a day of making responsible decisions as it would upon the automaton in the laboratory. At least we cannot assume, without investigation, that the effect would be the same. If we do not allow this assumption, however, the conclusions from such an experiment would have to be limited to such restricted statements that they would have little interest or practical value—*e.g.,* "Vitamin B has no effect upon fatigue in adding three-place numbers in columns of seven numbers for a period of four and a quarter hours."

It is quite possible that the major sources of fatigue in many sedentary and skilled types of work are not in the work itself so much as they are in the problems of social and emotional adjustment which the job entails. In the laboratory, these factors will operate in an entirely different manner from what they do in an office or factory. Laboratory studies would therefore give us an analysis of fatigue due primarily to the "work itself." Without indices of fatigue which could be applied "in the field," it would be impossible to assess the effects of these complex factors of adjustment upon the net fatigue.

Of course it is important to know the effects of both sets of factors—the work itself, and the other elements of the total situation. Laboratory studies, however, can give us only one side of the picture, unless it is possible to bring these emotional factors under laboratory control also. So far, this has not been achieved.

If it is known that a certain group of activities have similar properties from the point of view of fatigue, *e.g.,* a color-naming task and proofreading, and that one task is such that decrement is easily measurable, then the laboratory approach seems satisfactory and useful even if incomplete. This means that the use of the approach now under consideration depends upon extensive investigations of fatigue by other means which permit us to determine the relationships between the effects of different tasks. In other words, we must *first* show that color-naming in the laboratory has effects similar to those of proofreading in a publisher's office. Once these relationships are established, the method of decrement may be used as a short-cut procedure in solving a variety of practical problems.

2. Where a specific task is not amenable to study by the decrement method, it may be possible to measure performance periodically in some other task which is more readily measurable. This is known as a *fatigue test.* Thus, in order to study the effects of the job of

being a policeman, we might ask him to perform a series of tests in the morning, again before and after lunch, and finally as he leaves work. If his performance on these tests is diminished, we would assume that his job produces some kind of general fatigue which is reflected in other exacting activities. If some tests show greater effect than others, we would assume that there are specific effects of the work which are more limited in their nature—specific kinds of fatigue.

There are several major requirements for the use of the procedure. Most important, the tests must be sensitive to all important effects of the work. If not, a negative finding would be of no significance, since it might mean only that the tests missed the important effects of the work. This means, again, that the use of this procedure depends upon a fundamental analysis of the effects of the work by direct study, and that adequate fatigue tests come *after* this analysis rather than before.

In practice, this requirement has not been met. Instead, fatigue tests have been based upon an empirical search for activities in which performance varies inversely with the duration of work. The investigation is still in this stage, and has not yet reached the stage of inquiring whether these tests actually get at the important effects of the work.

3. A third possible way out of this difficulty is a more refined analysis of those tasks which do not appear under gross measurement of output to show decrement. There is evidence that fine control of co-ordination, accuracy of skilled activities, timing, and the like may deteriorate, even though the effect of fatigue is not clearly evident in the over-all rate of performance or accomplishment. It might be, for example, that more accurate measurement would indicate "blocking" or brief failures to respond even in such tasks. In skilled muscular movements, careful photographic records might reveal a slight decrease in accuracy of movement which is compensated for by increased effort. An investigation of techniques of this kind, which could be applied to a wide variety of tasks, would be of fundamental importance to the field of efficiency in human work.

4. A fourth method of solution of the problem may lie in physiological indices of effort and fatigue. (See Chapter 5.) The "involuntary" bodily activities such as circulation, muscular tonus, and blood chemistry have received considerable attention as possible indicators of effort in mental work as well as in heavier muscular effort. Some of these techniques, though still in a trial stage, show definite

promise of validity. Whether they will also be convenient enough for field studies still is undetermined. If they are not, it will be necessary to perform much of the work in the laboratory, hoping to extrapolate the conclusions to the everyday working situation.

Each of these approaches to the problem will be considered more thoroughly in the following sections, insofar as there is information available. Possibilities 2 and 4 have received more attention than the others, so that much of the discussion must of necessity be limited to them. This does not mean that these approaches necessarily show more promise. They have appeared more attractive to investigators so far, but they may receive relatively less attention as time goes on. The first approach mentioned is involved in many of the studies which will be described later when we take up such problems as distraction, the effects of alcohol, rest periods, and the like. There is little to be said now, however, upon the fundamental problems of the interrelations between fatigues developed under different conditions and with different tasks.[14] We therefore pass directly to a more detailed consideration of the second approach.

Tests of Fatigue

In the search for measurable activities which would be sensitive to the effects of prolonged mental work, and skilled activities of all kinds, a great variety of tasks have been tried out. In order to interpret the results of these varied experiments, we must keep in mind the precautions which are necessary to show that a given test is a suitable technique. Let us outline some of the main ones.

1. Statistical treatment of the results to determine whether the difference between before-work and after-work test scores is significant.

There are a number of factors which may lead to results which are so variable as to be indecisive:

(*a*) The test itself may not be a reliable measuring device—*e.g.*, it is not long enough to get a representative sample of performance at the time. This is a difficult factor to control, because lengthening the test may make the test itself so fatiguing as to cover up the effects of the fatigue which we are attempting to measure.

(*b*) The test may measure a function which fluctuates with a variety of other factors. That is, the test may be an accurate

[14] A. G. Bills, Transfer of fatigue and identical elements, *Journal of Experimental Psychology*, 1932, *15*, 23-36.

indicator of performance at the time of the test, but performance itself fluctuates widely with other uncontrolled factors.

(*c*) The use of a very small number of subjects, so that the results cannot be generalized beyond the particular subjects used in the experiments. If all the subjects consistently show the same trend, the number of subjects does not have to be very large. It should, however, be representative of a number of different sorts of individual—large, small; bright, dull; male, female; old, young; etc. If the results for each individual are extensive enough to be significant *for him,* and if several individuals of widely different kinds show similar trends, then the results are of value without carrying out a mass experiment.

N.B. If a test is to be used as a method of diagnosing fatigue in an individual, it must be shown to be significant for *individual diagnosis* as well as group differences. That is, a group of individuals might show a statistically significant trend if the group is very large, even though there are many individual exceptions and the changes in most individuals are very slight.

2. Adequate control periods to eliminate the effects of practice, and of variations with the time of day, independent of working activity.

For example, let us suppose that tested reaction time is lower after eight hours of office work than it was in the morning before work. We cannot conclude that this change is due to the work, even if it is not due to chance. The change might have occurred, even though the subject had not worked at all. The control tests taken on days when the subject does not work could determine whether this is possible.

On the other hand, the performance might be higher at night than it is in the morning, as a result of practice. If so, it is necessary to measure and to allow for this effect in order to determine what effect the work had. Or, preferably, the test is repeated so many times that the practice effect is finally eliminated as a factor.

3. Determination of the types of work for which a given test is applicable. For example, some tests might be affected primarily by postural restrictions in the work, others by the visual aspects of the task, and so on.[15] This means sampling a variety of tasks as well as a variety of individuals, repeating the precautions described above for each task.

The experimenters who have studied the problem of fatigue by such tests have given considerable attention to adequate control. As

[15] See T. A. Ryan, Varieties of fatigue, *American Journal of Psychology,* 1944, *57,* 565-569.

in the studies of muscular effort, however, there has been widespread neglect of proper statistical procedures. The reliability of the tests has often been completely unknown, or has been deduced from logical considerations rather than from a direct analysis. This would not have been fatal if there had been an over-all analysis of the statistical significance of the test variations in the experiment. Too often, however, not even this was done.

For example, Kraepelin, his students, and co-workers have provided us with a number of studies of fatigue-tests in mental work. Yet some of these studies fail to report even the number of subjects employed or the number of observations on each subject, to say nothing of the data upon variability of the results. Conclusions are drawn upon the basis of simple averages.

It is true, of course, that an experienced researcher in any field can tell by his direct observations whether or not the methods have great variability. He is still liable to error, however, and he should always check his impressions by statistical analysis of the data. If he does so, we assume that he will report the fact along with his results so that he can demonstrate clearly the value of his results. Unfortunately, however, many men who early received their training in this field have failed to realize this need.

Since so few of the experiments on fatigue tests so far reported have met the requirements laid down in the preceding discussion, we shall not devote much space to them except to indicate the general approaches that have been taken.[16] The proposed tests may be roughly classified under the following headings:

1. Tests of functional components.
   - (*a*) Muscular strength and control (ergograph, tapping).
   - (*b*) Sensory acuity and discrimination, *e.g.,* discrimination of two points upon the skin.
   - (*c*) Reaction time tests of varying degrees of complexity.
   - (*d*) Attention and concentration.
   - (*e*) Memory.

2. Tests involving a variety of components (sometimes closely related to the fatiguing work itself).

These cannot be classified readily, but a few examples would include these: addition or other calculation, writing to dictation, completing sentences, tests of speed of number checking.

---

[16] For more detailed accounts of these tests and results see M. Offner's *Mental Fatigue,* 1911, or *Die geistige Ermüdung,* 1928. Also, O. Graf's *Über Ermüdungsmessung durch Stichproben,* Kraepelin's *Psychologische Arbeiten,* 1925, *8,* 1-16.

For many of these tests, the lack of statistical control of a single investigator has been brought to light by the attempts of other experimenters to duplicate the results. Thus the literature is full of claims and counterclaims as to the value of a particular test. This could have been avoided if the precautions we mention had been taken. Offner, who summarized the work up to 1928, based his evaluations to a great extent upon these contradictions.[17] From this rather haphazard collection of information, Offner concludes that muscular and simple sensory tests have little value in measuring the effects of "mental" work. Greatest hope lies in the more complex types listed above under (2).

Examples of Fatigue Test Studies.—Here are two examples of the use of the fatigue-test approach. The first was concerned with the effects of loss of sleep.[18] After observing her normal performance in the tests to be described, the experimenter limited her sleep for three successive nights to 1½, 3½, and 5½ hours respectively. The tests were administered each day during the deprivation and for several days thereafter. Sometimes there were two sleepless periods separated by several normal days, with the test records complete for each day throughout the whole span.

The tests were as follows:

1. Dotting test. Small red circles on a moving tape had to be dotted as they passed. Circles of other colors had to be skipped, while the subject tapped a key with the other hand each time one of these circles of different color appeared.
2. Association test. A series of words with logical connections between them was read once. The subject tried to reproduce the list in the correct sequence immediately after it was read.
3. Illusion of reversible perspective. This test involved a figure which was ambiguous so that it would be seen in either of two ways. The subject observed the figure for a period of time, signaling whenever the perspective was reversed.
4. Speed of tapping.
5. Learning and relearning of nonsense syllables.

The most marked changes appeared in the first two of the tests. The performance was normal or above on days in which the loss of

[17] He also discards certain tests which he feels are too complex or difficult to administer to large groups, even if there is some hope that they may have value otherwise. This is due primarily to his interest in the use of fatigue tests in the schools for controlling the hours of work, scheduling, and the like.

[18] May Smith, A contribution to the study of fatigue, *British Journal of Psychology,* 1916, *8,* 327-350.

sleep occurred.[19] It went into a decline after that, however, and sometimes remained below normal for more than a week. If further sleep was lost during this subnormal period, the decrease in sleep had the effect of returning the performance to normal. These results are, of course, only illustrative of the method. There was only one

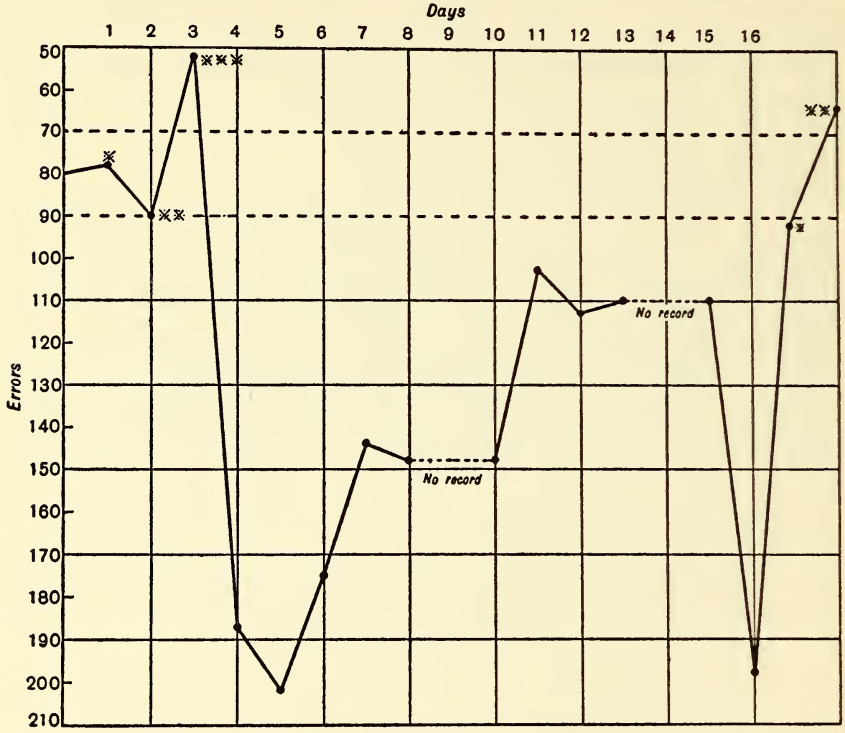


Figure 10. Test Performance Following Three Nights of Diminished Sleep (Dotting Test)

* After 1st vigil—1½ hours of sleep.
** After 2nd vigil—3½ hours of sleep.
*** After 3rd vigil—5½ hours of sleep.
(Vigils were successive nights of diminished sleep.)

(From Smith, *op. cit.*, p. 333.)

subject, so that the value of the results is indeterminate, even though the changes were regular enough to be statistically significant for that subject.

Another example is an experiment upon the effects of automobile driving.[20] This is a more recent study, more carefully controlled.

[19] Smith, *op. cit.*, pp. 333, 335.

[20] A. H. Ryan and Mary Warner, The effect of automobile driving on the reactions of the driver, *American Journal of Psychology*, 1936, *48*, 403-421.

The operators drove a distance of about 300 miles on "driving days," which came every other day. On the "control" days in between, they drove for an hour before the first test of the day, just as they did on the experimental days. The rest of the day was spent, however, in light activity and recreation. Six drivers were studied for a total of

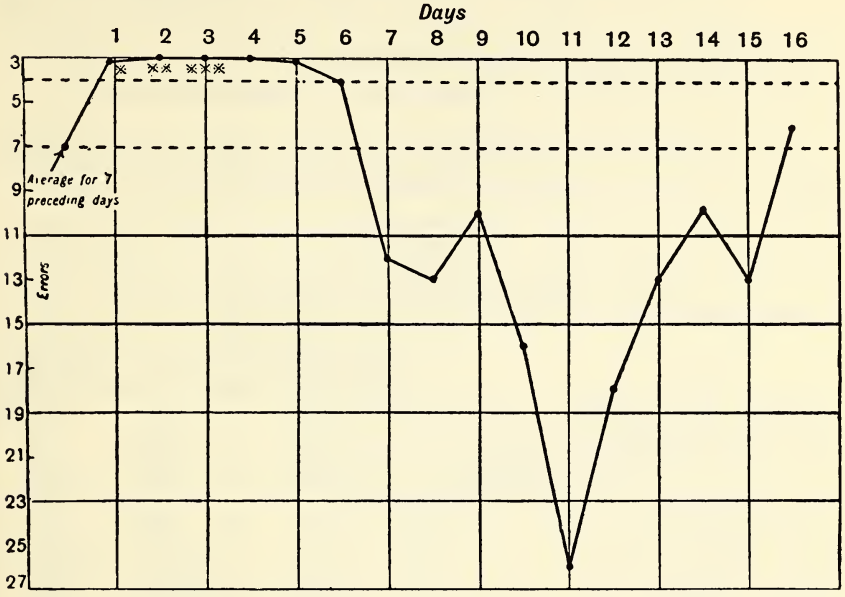


Figure 11. Test Performance Following Three Nights of Diminished Sleep (Associated Words)

* After 1st vigil—1½ hours of sleep.
** After 2nd vigil—3½ hours of sleep.
*** After 3rd vigil—5½ hours of sleep.
(Vigils were successive nights of diminished sleep.)

(From Smith, *op. cit.*, p. 335.)

120 driving days. The results are given in terms of averages for the six drivers.

The tests, which were repeated each day in the morning and again in the evening after driving or light activity, were as follows:

1. Vascular skin reaction, a physiological test designed by A. H. Ryan, measuring the time of disappearance of a red streak on the skin produced by a stroke of a blunt instrument.
2. Postural steadiness, the amount of sway in standing with the eyes closed.
3. Hand-eye co-ordination, inserting a metal stylus into holes of various sizes at a fixed rate, trying not to touch the edges.

4. "Visual efficiency," Ferree's test, measuring the length of time a subject can maintain focus upon the letters *li*. The subject signals whenever the letters become blurred. (See Chapter 6 for a discussion of this type of test.)
5. Color naming, measuring the time required and the errors made in naming a series of 1200 color squares. A test which has been used in a number of special investigations—*e.g.*, the effects of alcohol upon psychological functions.
6. Mental addition. Adding columns of fifteen digits each.

The results of the comparison of morning and evening tests are shown in Table 8, taken from Ryan and Warner's report.

TABLE 8*

CHANGE IN PERFORMANCE FROM MORNING TO EVENING

Tests	Percentage of Change: Driving Days	Percentage of Change: Control Days	Diff. / P.E. Diff.
Vascular Skin Reaction	— 12.38	— 1.21	12.93**
Postural Steadiness	8.32	.20	3.83
Hand-eye Co-ordination: Form A	— 2.92	.39	3.64
Form B†	23.58	— 9.78	7.76
Visual Efficiency	— 2.62	— .80	6.62
Color Naming: Time	3.93	— 1.65	8.92
Color Naming: Errors	46.79	— 16.28	12.45
Mental Addition	8.46	— 3.59	8.31

* Modified from Ryan and Warner, *op. cit.*, p. 409.

** A ratio of difference to probable error of difference should usually be 4 or greater if the difference is to be considered significant. Values slightly smaller are sometimes considered as suggestive.

† Form B of the co-ordination test was longer and more difficult than Form A.

What conclusions can be drawn from a study of this kind? Ryan and Warner, authors of the report, draw the following:

"(1) There is a demonstrable fatigue effect of a long automobile drive on body reactions.

(2) Long automobile drives tend to decrease the fading time of the vascular skin reaction; increase unsteadiness in standing; decrease the accuracy of hand-eye co-ordination; decrease the visual efficiency; decrease the speed and accuracy of mental addition.

(3) Greater variability occurs within the test scores made after long automobile drives than before them in postural steadiness, hand-eye co-ordination, and mental addition.

(4) The tendency of long automobile drives is to produce a loss of effectiveness of certain sensory discriminations, association processes, and motor reactions similar to those required in driving. These observations suggest that the effect of a long automobile drive may render a driver temporarily prone to accidents." [21]

These conclusions are cautious, and they appear justified by the results we have reported. Since accident liability in performance cannot be easily measured, it seems necessary to study other functions which lend themselves more readily to reliable measurement.

Since, however, we are concerned with the evaluation of this general approach to the measurement of fatigue, we must examine the meaning of this study more fully. We must raise some questions which refer not only to the particular problem of this experiment, but also to any other problems which might be approached with the same general method of fatigue tests.

TABLE 9*

Proportion of Days Showing Decreased Performance

Test	Driving Days	*N*	Control Days	*N*
Vascular Skin Reaction	100%	120	52%	23
Steadiness (sway)	65	"	46	28
Hand-eye Co-ordination: A	54	"	32	"
B	65	"	32	"
Visual Efficiency	89	100	74	23
Color Naming: Time	75	120	36	28
Color Naming: Errors	77	"	21	"
Mental Addition	68	"	21	"

* Assembled from data in Ryan and Warner, *op. cit.*

1. In order to obtain reliable measurements, the time spent upon each subject was so great that only six subjects could be studied. Thus, while the results are evidently significant for the group, there is still the question of how the driving would affect other groups of individuals. Since no results for single individuals are given, we cannot tell how much of the variability of results is due to the variations of each individual and how much to differences between individuals independent of their temporary variations. If each individual was consistent but there was wide variation between individuals, there is a definite need to study more subjects. Table 9 shows some

[21] Ryan and Warner, *op. cit.*, 420-421.

data which Ryan and Warner give as an indication of the consistency of the results, but these values are also composite results for several individuals.

2. The decision that these tests represent functions which are involved in accident liability is based upon judgment, and it is doubtful whether it could be directly proved. Such judgments are very difficult to make, as we shall see when we discuss accident control. (See Chapter 12.) They must be made with great caution.[22]

3. How are we to measure the *amount* of fatigue produced by the driving? A number of important practical applications would depend upon the answer to this question. How many hours of driving can be performed safely before the changes in function become great enough to be dangerous? Suppose that a new car design is to be tested for ease of driving. Can we differentiate between the effects of driving two different cars by a procedure like that used by Ryan and Warner?

All we know is that the changes reported are significant statistically. How important they are for practical driving is not determined. Also, it is doubtful if the tests are sensitive enough to distinguish between the effects of driving under two different conditions. At least these are questions which are still to be determined.

It is likely that further research into the problem of fatigue tests will be valuable in throwing more light upon the interrelations of the effects of various kinds of work. That they will provide us with a method of measuring or even estimating the cost of work at any time in the near future is highly doubtful, for the reasons which we discussed in evaluating A. H. Ryan's study.

There is probably much more to be gained from this type of research if it is made more systematic, and if it is used as a device for finding out the nature of the effects of work (fatigue). Those fatigue tests which have measured so-called "elementary" functions have been based upon extremely vague hypotheses that fatigue affects "attention," "speed of reaction," or similar broad aspects of functioning. It has then been assumed that the tests employed are valid measures of these functional characteristics. Those who have regarded the fatigue test as a measure of elements or parts of the job itself under controlled conditions (as in A. H. Ryan's study of drivers) have not been concerned with problems of the general effects of fatigue. In fact, studies of this latter kind are simply substitutes

[22] See H. M. Johnson, Index-numerology and measures of impairment, *American Journal of Psychology*, 1943, *56*, 551-558.

for direct measurement of work decrement. They gain, in that the tests are more amenable to quantification than the task itself, but they lose, in that the validity of the tests as samples of performance in the task itself is unknown and usually must be assumed.

These studies are and will be more enlightening if we take them out of the context of measurement of fatigue and fatigue "tests" altogether. If, instead, we ask what kinds of activity are functionally so related that work in one task produces decrement in the others, we are entering the context of the general problem of fatigue *transfer* rather than of fatigue *measurement*. More information of this kind, on a wider variety of activities, could lead to the next step, which would be the formulation of general principles to cover these functional relationships. We might then be on our way to discovery of just what it is which brings about the functional relationships between certain tasks—what bodily resources they draw upon in common, what common elements there are in a group of related activities, and so on.[23] Such a program calls also for the use of many physiological techniques that are to be discussed in the next section.

Some attempts have already been made in this direction, but they have been based upon such scanty information that the principles enunciated have been, of necessity, exceedingly vague and trite.

Bills, for example, summarizes the conditions "under which the decrement developed in one task can be expected to transfer to a second task which is engaged in directly afterward. These conditions are:

> When the tasks are both mental, or both physical, or when both have elements of each.
>
> When the two tasks are quite similar, involving the use of some of the same neuromuscular processes.
>
> When the work set, *i.e.*, the directing and controlling attentive set, is common to both.
>
> When the first task is so complex that it results in the widespread fatigue of many mechanisms.
>
> When the setting in which the work is done remains unchanged, provided this feature was the fatiguing element in the first task." [24]

While this conclusion of Bills is all that is justified by our present knowledge of the transfer of fatigue, it is obvious that it does not

---

[23] As an example of this kind of analysis applied to visual work, see M. E. Bitterman, Transfer of decrement in ocular tasks, *American Journal of Psychology*, 1946, *59*, 422-438.

[24] Bills, *op. cit.*, p. 111.

carry us very far toward an understanding of fatigue. Of course, fatigue-transfer is a problem of interest in itself, but it has even greater potential importance if it can be used to throw light upon the underlying bodily events responsible for fatigue and for transfer of decrement.

A fact which introduces some confusion into the field of fatigue testing is that sometimes work produces an *improvement* in test scores rather than an impairment. If the improvement is really due to the work and not to some other factor, it may be interpreted in several ways. First, the effects of a specific kind of fatigue may be to reduce capacity for some tasks (closely related to the fatiguing task) while increasing the capacity for unrelated tasks. It might be simply that the task provides a change or relief from the restraints imposed by the fatiguing work. Another interpretation is suggested by Johnson, who compares fatigue with the effects of alcoholic intoxication. The apparent improvement may be only temporary, and may not represent a real benefit to the organism.[25] Instead, it may indicate a condition which is ultimately harmful because it makes additional demands upon the reserves of the body.

Perhaps the best way to approach this paradoxical improvement resulting from work is to wait until we learn what kinds of bodily change underlie the effect. "Improvement" and "impairment" are values arbitrarily imposed upon the test scores in relation to their involvement in practical affairs. It is as if we were to call an increase in heart rate an "improvement," and a decrease an "impairment." Obviously, we are justified only in saying that the heart rate has changed. In the same way, a change in test performance is to be regarded as a result of complex internal bodily changes. Whether or not these changes represent important impairment of bodily economy still remains to be determined.

So far, what has happened is that those tests which fail to show decrement are discarded as failing to reflect the effects of fatigue. As a result, those tests which show improvement are given the same treatment as those tests which show no effect at all. If, as suggested above, the tests were used to examine the various effects of work without pointing so specifically toward developing indices or measures of fatigue, any effects of work would be important, no matter what direction of effect happens to be involved. The task of the scientist is, then, to piece all this information together so as to formulate hypotheses concerning the underlying bodily changes.

---

[25] H. M. Johnson, The real meaning of fatigue, *Harper's Magazine*, 1929, Vol. 158, 186-193.

Change of Capacity as Indicating Change in Efficiency.—Another way in which tests of capacity are often used in the study of efficiency differs from the approach we have spoken of as the "fatigue test." In investigating the effects of drugs, changes in atmospheric conditions, or other factors, the experimenter may measure the psychological effects by comparing test performance with and without the drug, or under varying temperatures, oxygen pressures, or the like. If it is found that alcohol, for example, impairs ability to type, it is usually said that alcohol impairs "efficiency" in this performance.

In some cases, this mode of expression may be no more than a confusion of output alone with efficiency, a usage which we criticized at the outset of this book. Frequently, however, efficiency as a ratio of output and cost seems to be implied. We must therefore consider whether this conclusion about efficiency on the basis of impaired capacity is justified.

Let us accept, at least for purposes of argument, the position that the tests of reaction time, number checking, color naming, speed of adding or learning, which are used in these experiments, do measure capacity. At least the subjects are working under high motivation, since they are usually co-operative and told to do their best.

If we grant that the subjects are working at their limit of performance in the test, and that the drug under study is found to produce a reduction in this limit, we still have to relate this finding to the concept of efficiency. The argument would have to be something as follows:

1. As the organism approaches its limits of performance, it becomes less efficient. For example, a sprinter at top speed uses more oxygen per mile than he does at an easy lope or at a brisk walk.
2. If the worker's limit of performance (capacity) is reduced by some factor, such as a drug, his usual rate of performance on his job will be closer to the limit than it normally is.
3. Although his rate of performance in his normal tasks does not change, he is now exerting greater effort, just as though he had speeded up in his rate of work. (See definition of *effort* in Chapter 2, page 22.)
4. Therefore the drug has not only reduced the maximum limits of performance, but it has also reduced the efficiency with which the task can be carried on at normal rates.

Notice the assumptions involved. First it is assumed that these tests measure capacity, or at least that the change under the influence of the drug indicates a change in capacity. Then it is assumed that

efficiency of performance at a given rate is a function, not of the rate itself, but of the relation of the rate to maximum possible performance (*effort*). In other words, we are assuming that other elements of cost increase in proportion to the effort involved.

Another way in which we might justify using these results as indices of efficiency would be as follows: It is assumed that these tests measure aspects of alertness, speed of reaction, and attention, all of which are necessary for accurate performance of complex and dangerous jobs such as automobile driving or machine operation. If these capacities are impaired, the danger of serious error and accident is increased, an obvious source of inefficiency.

At present this latter set of assumptions is not so well founded as the first. There is no evidence for the validity of the tests commonly used in experimental studies of drugs or physical factors as measures of the *kind* of alertness or reaction speed involved in dangerous occupations. From what is known, the writer would prefer the former approach in terms of the relationship of performance to capacity.

Let us return now to the assumption which we accepted for purposes of argument, namely, that test performance reflects changes in capacity. Essentially the same question arises in any measure of the effects of some factor upon performance. Thus the measurement of fatigue in terms of output decrement assumes that decreased capacity, due to fatigue, is reflected in reduced output. The measurement of performance in a fatigue "test" before and after work is based upon a similar assumption. As we have said (see Chapter 2), capacity is the maximum performance of which a subject is capable under given conditions. Even though the motivation of the subjects of these experiments is high, it is probably not maximal. In fact, the term "capacity" refers to a limiting value which is never reached in practice; it is only approached.

The crucial question, however, is not whether test performance is nearly maximal, but whether the *changes* in performance which occur under experimental variation accurately reflect the changing limits. There is some evidence that even under laboratory conditions the worker has enough reserve so that when conditions become adverse he exerts greater effort, and thus partially or wholly obliterates the effect upon performance which we should expect to appear. (This statement is based upon indirect evidence, or upon physiological measures to be discussed on subsequent pages.)

Thus, even under carefully controlled conditions with highly motivated subjects, changes in test performance are only partially a reflection of changes in capacity. It is probable that a *reduced* per-

formance under these conditions *does* indicate a *reduced* capacity. A failure to observe decreased performance would not, however, be a conclusive demonstration that the capacity remains the same.

Summary of Fatigue as Indicated by Decrement and Fatigue Tests.—Our analysis of the studies of output decrement in sedentary activities has led us to the conclusion that this index has little value as a reflection of fatigue in ordinary daily activity. Outside of the laboratory there are many factors which prevent any change in capacity of the worker from showing its effect in a decline of actual productive rate. Even in the laboratory it is only specially selected tasks which show clear-cut and reliable decrement which is a function of the length of working period.

As in the case of muscular work, then, it is possible to study output decrement in the laboratory if the experimenter can limit himself to certain narrow restrictions in the experimental task and in the conditions of work. In most cases, however, these restrictions mean that the conclusions apply to such an unusual kind of task that the application of these conclusions outside of the laboratory is of little value. Only if the laboratory task respresents a phase or component of some more complex normal task is this approach likely to be useful.

Among the other possible approaches to the study of fatigue, one other general method has been given extended consideration in this chapter. This is the procedure called the "fatigue test." We have described some of the normal research precautions which must be applied before the results of such a study can be interpreted at all. This discussion of methodology was essential because it has been neglected in so many of the published researches in this field. The result is that only a few researches upon "fatigue tests," can be given serious consideration. The remainder contribute only results which may be spurious because of the unreliability of the tests, insufficiency of sampling, or uncontrolled variables.

When the ordinary experimental and statistical precautions have been taken, the approach by way of fatigue tests is a way of gathering valuable information upon the effects of work. The general scientific program based upon fatigue tests would be to determine the interrelations of a great variety of tasks from the point of view of the transfer of fatigue-effects. Such a program would be directed toward finding the underlying effects of work by a study of the pattern of interrelationships brought out from fatigue tests.

The use of fatigue tests for the solution of technological problems requires other steps beyond the general scientific program sketched in

the preceding paragraph. In evaluating the results of sample fatigue-testing researches, we have found certain deficiencies in the results from the point of view of application. Since we treated these difficulties in piecemeal fashion, as the problems arose in discussing our examples, we must now summarize what is necessary in order to make fatigue tests practically useful.

1. Most studies of fatigue tests have been content with showing that work produces a significant change in test performance—a change which is significantly greater than those which appear in the absence of work. We must also show that the test performance varies with the amount of work. In other words, we wish to show that the test is sensitive to varying amounts of effort, and the only way open (in the absence of a direct measure of effort) is to vary the total amount of preceding work over a considerable range.

If a test has met this requirement, it will be useful for certain problems of machine design or work-methods. In Ryan and Warner's methods, for example, if we could show that six hours of driving car A is the equivalent of eight hours of driving car B, then we should have weighty evidence in favor of the design of car A. Until fatigue tests are made sensitive enough to produce results like this, however, the method remains of largely theoretical importance.

2. It is also necessary to evaluate the changes measured by the fatigue test in terms of their meaning for the total economy of the individual. The fact that car driving slows verbal reaction time (color naming), for example, is meaningless in itself. To evaluate this change in speed as a deterioration of the individual is to assume either (1) that slowing of reaction means that the individual is less capable of performing his job or the other tasks which he is expected to carry out, or (2) the slowing of reaction is a reflection of some physiological change which is harmful to the individual if it exceeds certain limits. Neither of these assumptions has any real justification in terms of experimental evidence for the reaction time test or for any of the other tests in use in the studies we have sampled. Of course, it seems only "common sense" to make the assumptions. As we have stressed before, however, this does not make the assumptions necessarily correct, and experimental justification is still needed.

The program of fatigue-testing for purposes of application to industrial problems must therefore be a very extensive one. It is not only necessary to investigate the effects of work upon the tests, but we must

know how performance on these tests is related to other criteria of the capability, proneness to accident, health, or general satisfaction of the individual.

3. For some purposes, fatigue tests must be applied upon an individual basis. In these cases, the tests which are to be applied must meet further requirements. It must be shown that the test is significantly affected by work in the single individual, as well as in the large group of subjects. It would be quite possible that a test could show a significant *average effect* in a large group of subjects, even though some of the subjects consistently showed the reverse effect or no effect at all. A test which did not show such wide individual variations would be valuable for both individual and group studies, since variability in the latter case would be reduced by the elimination of the extreme individual variation.

4. A program of fatigue-testing must sample a wide variety of tasks and activities—as nearly a complete sampling of human activities as possible. To return to an example used once before, to show that driving car A has a greater effect upon a battery of fatigue tests than that produced by car B, is not conclusive unless we rule out the possibility that another battery of tests could show a reversed result. This requirement means that we must have a wide sample of test activities, as well as a sufficient sample of activity upon each test. Once it is shown that the results for certain groups of tests are consistent with one another, we can save time by using only one sample from the related group of tests. Before this can be done, however, we must again await the results of more extended research.

It is thus evident that, although fatigue-testing is a field of research which must be carried on for some time, it is of a sort which lays the groundwork for later application. The applications can be made only when and if the above characteristics of fatigue tests can be demonstrated, which means only after much intensive additional investigation.

The literature upon fatigue-testing in recent years gives the impression that interest in the problem has been declining. Perhaps this is due to discouragement at the vastness of the problem, or to the belief that so much has already been attempted with so few tangible results. Yet the amount of satisfactorily controlled work has been relatively small in relation to the size of the problem. Perhaps progress can be accelerated if new procedures of testing, not now envisaged, can be discovered.

In all tests based upon psychological performance, however, there will always be a difficulty of controlling the motivation of the subjects taking the tests. For this reason, as well as for others, many researchers have preferred to search for indices of fatigue and effort which require only the measurement of physiological changes. These techniques will be the subject of the following chapter.

Chapter 5

MEASURING THE COST OF WORK: III. PHYSIOLOGICAL TESTS OF EFFORT AND FATIGUE IN SEDENTARY WORK; STATISTICAL INDICES OF EFFICIENCY

Since both sedentary and muscular work entail the co-operative functioning of many parts of the organism, it is reasonable to expect that sedentary work would be accompanied by various measurable physiological changes, and that fatigue resulting from sedentary work would have a physiological basis. A large body of experimental data has therefore been assembled in the search for physiological correlates of mental effort and the fatigue resulting from mental work.

Although we have already given some preliminary consideration to the concept of effort in an introductory chapter (Chapter 2), it should be enlightening to consider this concept more thoroughly at this point. If we seek correlates and indices of effort, we must first have a clear notion of what effort is.

An Analysis of the Concept of Effort

Although the term *effort* plays a crucial part in many phases of applied psychology and industrial engineering, it is a term whose meaning has been very inadequately defined. Time standards for jobs are supposed to be determined in such a way that workers of equal training and experience would receive equal pay for equal effort. Methods-design is the problem of finding less effortful methods of performing a job. The term is used in common-sense analyses of these problems without definition, and a need for definition is not felt because it refers to a meaningful aspect of common experience.

The experimental study of efficiency, however, requires that we deal with definable and, if possible, quantifiable phenomena. Often, terms from common speech cover too many distinct and unrelated phenomena to be scientifically useful, and such terms must be more carefully limited in their use if they are to be valuable to the research worker. We must therefore consider the various meanings of the term, in order to find that meaning which will make it most readily

amenable to experimental study, and, more important, which will delimit a unique and important area for investigation.

The possibilities which we shall seek to weigh are as follows:

(*a*) Effort is equivalent to energy consumption.

(*b*) Effort is a general term covering a number of separate elements of the cost of work.

(*c*) Effort is a term describing certain aspects of psychological functioning—the pushing on toward completion of a task. It is a term describing the "experience" of the worker as he performs his job.

(*d*) Effort refers to the rate of performance of an individual in relation to the maximum possible rate of performance at the time and under the given conditions.

Each of these notions of effort has some sanction in common usage, and each leads to different consequences in practice. The first two meanings would imply that effort is simply a synonym for other aspects of the cost of work, so that the word could be used for variety, or for brevity, as in the case of (*b*). Unfortunately, common sense implies that there is more involved, and we must investigate this possibility that effort concerns an aspect of activity not covered by other terms. We do not desire to overlook any possibilities even if they are complicated or inconvenient.

(*a*) Certainly, effort is more than energy in the strictly physical usage of the latter term. As we have seen, mental work which the worker would describe as extremely effortful does not show a corresponding increase in metabolic activity in the organism. In reference to possibility (*b*), all the elements of cost so far considered, such as fatigue, dissatisfaction, loss of health, risk of accident, refer to effects of the work upon the individual, for the most part, *aftereffects*. Effort, however, is supposed to refer to some aspect of the activity as it progresses—it is a characteristic of the work-in-course. These other elements of the cost of work cannot, therefore, be identified with effort, although we may expect a close correlation between total cost, as measured by them, and the level and duration of effort, however effort is defined.

(*c*) The third possible meaning would be stigmatized as "introspective" by those psychologists who insist upon being "objective" at all costs. The present writer would have no prejudice against such an approach if it could be made workable. In fact, we may eventually be forced to adopt this approach to effort, since it is the closest to the

common-sense meaning, and therefore is intimately involved in the practical problems which are placed before the psychologist in industry. That is, if the worker or the industrialist seeks to have pay awarded on the basis of "equal pay for equal effort," the thing which is sought is only indirectly indicated by fatigue and loss of health. What interests the practical man is that a person who is "working hard," "who is exerting himself," or "taking pains," should be paid accordingly.

If fatigue or other subsequent indicators were easily and accurately measurable, then the problem might be solved as a matter of expediency by equating effort with these effects or correlates of effort. As we have seen in the preceding chapters, however, fatigue is a difficult problem in itself. Perhaps suitable direct criteria of effort would be more practically useful than measures of fatigue. Certainly they would be more immediately measurable if they could be discovered, since we would have a measure which could be applied during the course of the work itself.

Effort could be defined in terms of Bentley's functional approach to psychological description.[1] Lacking experimental studies directed toward obtaining refined descriptions of the effortful aspects of activity, we are limited to conjectures as to the probable nature of this description.[2] Effort obviously refers to the *determined* or *searching* aspects of function.

Such a descriptive study should ultimately be carried out, both because it would contribute to our fundamental understanding of the organism at work, and also because it might suggest new modes of approach to the practical problems of measuring and controlling effort. At present, however, one cannot see what such a description would contribute. Moreover, the descriptive approach still leaves open the problem of measurement, which is so important in the solution of industrial problems.

(*d*) Although definite advances in the status of our problem may come from the descriptive approach just described, nevertheless we need some simpler and more definite statement of what is meant by effort. We have found the definition given above under (*d*) a useful way of conceiving effort, since it is in terms which are at least approximately measurable, and also remains fairly close to the common-

---

[1] M. Bentley, *The New Field of Psychology,* New York, D. Appleton-Century Co., Inc., 1934.

[2] For illustrations of the technique employed in such a functional description see the series of studies published as the Cornell Studies in Dynasomatic Psychology, under Bentley's editorship, *American Journal of Psychology,* 1938, *51,* 203-360.

sense meaning of the term. It would appear to be the meaning which is implicitly involved in many of the physiological studies of mental work which we are now about to examine.

In spite of our adoption of definition (*d*) as being the most manageable statement of the meaning of effort, it must be recognized that there are definite flaws involved. One of these is the practical difficulty of measuring capacity as of a given moment. This difficulty can probably be overcome by statistical and experimental controls. Another flaw is inherent in the use of a rate of output as a measure. Suppose, for example, that the capacity of a given individual at a given time is 100 units per hour. He increases his rate of work from 50 to 60 units per hour, an increase of 10 percentage units of effort. Later, he increases his rate once more from 60 to 70 units per hour, or by another equal step upon the arbitrary scale we have set up. Common observation, however, would make us suspect that the second increase is more difficult than the first, that the change in effort is greater in the second step than in the first. This would be an especially important factor when the performance varies close to the upper limit.

Lacking any way of measuring effort in anything except output terms, however, we must be content with it for the time being. We must be cautious, though, about assuming that effort is to be defined as a *linear* function of the ratio of performance to capacity. More complete knowledge may lead us to adopt a definite functional relationship. At present we must be content with the relationships "more" and "less." This is at least a first step toward measurement. In later discussions in this book we shall suggest researches which are needed in order to throw more light upon effort and its relation to other factors in work.

The extended discussion of tests of fatigue involving psychological activities included many considerations which apply equally well to physiological measures of effort and fatigue. Essentially the same requirements of experimental and statistical control apply to these physiological tests, so it will be necessary to discuss only the extent to which these requirements have been met.

A variety of physiological changes in the state of the organism during and after prolonged mental work have been reported. Changes in pulse rate and blood pressure, breathing, the vascular state of the skin (cf. A. H. Ryan's test mentioned above), changes in cell bodies and chemistry of the blood, urinary changes, muscular tonus or tension, have all been reported by some investigators to be correlated with the length or difficulty of mental work. Of these, the major

emphasis has been placed upon circulatory measures (pulse and blood pressure) and measures of muscular tension.

Most of these functions are known to be closely related to heavy muscular effort. They are also activities of bodily systems which are involved in emotional seizures. It has been argued that they should therefore be related to effort in general, and that they might serve as measures of effort even though metabolic indices are unreliable when applied to mental tasks.

It has also been claimed that "thinking is muscular work" [3] because all psychological activities seem to be accompanied by muscular contractions, and also because changes in muscle tension may facilitate or hinder psychological performance. The bodily phenomena accompanying tasks involving alertness and attention should therefore reflect patterns similar to those involved in muscular work. Such *a priori* arguments are, however, inconclusive for several reasons.

In the first place, the argument that thinking *is* muscular work, and that attention *is* muscular tension, goes far beyond the demonstrated facts. It is true that we usually find muscular tension increasing when the subject is alert, when he is solving a problem in mental arithmetic, and so on. This does not justify equating the two types of activity, and even though we would grant that tension is an element in all activity, "mental" as well as muscular, mental work is certainly much more than a pattern of taut muscles. Even in work which is classified as primarily muscular, it has not been demonstrated that all fatigue is localized in the muscular apparatus. Particularly in skilled muscular activities, it is probable that effort is involved in the *control* of the muscular patterns as well as in the contractions themselves.

The above objection still admits a role of muscular tension in mental work. In fact, we shall attribute considerable importance to this role, in a somewhat different way. There is, however, still another flaw in the *a priori* argument that physiological patterns should be similar in the whole range of mental and muscular work. In those activities in which the muscular activity is not central and overt—*e.g.,* mental arithmetic, composing a report, and the like—the muscular tensions are present but they are so slight as to call for relatively little energy. If they were not, this energy could be clearly demonstrated by the metabolic technique.

---

[3] Bills, *op. cit.,* pp. 10-13. See also, A. G. Bills, Fatigue in mental work, *Physiological Review,* 1937, *17,* 436-453.

How, then, can it be argued that fatigue in mental work is similar to the fatigue due to heavy muscular labor? If it were, we should be able to continue our efforts in sedentary tasks almost indefinitely without appreciable fatigue, so long as we changed our position often enough to avoid muscular stiffness.

Perhaps some will say that the "real" fatigue in mental work is relatively slight or non-existent; that the fatigue which is reported after several hours of work upon a difficult problem is only "subjective." Some authors, indeed, seem to imply that reported fatigue is a sort of unreal existence without organic basis. "Subjective" fatigue must have *some* bodily basis, however, and the finding that it is not the same basis as muscular fatigue would not make it belong to the realm of "purely mental existence" (whatever that may be).

Others may say that the primary source of fatigue in mental tasks is due to emotional strain rather than the effort involved in just paying attention. This is very probable, but the fatigue resulting from emotional seizure cannot be accounted for on the basis of muscular tensions either. These tensions are likewise not great enough to explain the severe effects of a prolonged emotional upset purely in terms of the energy consumed.

Naturally, the value of physiological indices of circulation or tension can be settled only by direct experimentation. The *a priori* arguments analyzed above are mentioned because they are so frequently brought up in discussions of fatigue, not because they contribute to the solution of the problem. In fact they are more likely to lead us astray in any search for a solution.

The factual evidence on the effect of various kinds of sedentary work upon physiological functions is very similar to the evidence upon psychological tests of fatigue. There is the same tendency to report results of a small group of subjects, without a statistical analysis of the data which are reported. Again the result is a series of conflicting reports upon the same physiological indicator. For example, the pulse rate is variously reported as increasing, decreasing, or showing no change while a subject works upon numerical computation or similar tasks.

Circulatory Indices.—Bills concludes that all work involves an increased pulse rate, but that sedentary work may appear to reduce the rate if it follows soon after walking or other overt muscular activity.[4] This is because the level of circulatory activity in sedentary work is

[4] A. G. Bills, Fatigue in mental work, *Physiological Review*, 1937, *17*, 436-453.

lower than that required in muscular activity. In order to demonstrate clearly the effects upon circulation, it is necessary to establish a base line for the subject by an extended rest before the work begins. This is true, of course, of any physiological indicator which may be used.

If the work is done calmly with little emotional strain, pulse, blood pressure, and other circulatory indices seem to reflect the relative amount of muscular activity going on in the body. As such, they can be used as rough measures of metabolic cost of work. They are somewhat easier to apply than direct measures of metabolism, however, and for that reason may be more useful under many conditions. Special emotional strain produced by the working situation will also be reflected in these circulatory measures, unless the subject was also upset during the period of rest which established the base line. If the situation is such that it produces an unfavorable anticipation *before* the work begins, there is no way of establishing a base line and no way of estimating circulatory increases.

It is probable that, under carefully controlled conditions, work such as problem solving or mental calculation and the like will produce significant increases in pulse rate and blood pressure over the resting values.

Gillespie, for example, found that both pulse rate and blood pressure were higher in subjects doing mental calculation than they were when the subjects merely read the numbers.[5] No statistical treatment is given, however, and there were some exceptions to the general trends. He concludes that the rise in blood pressure is not proportional to the difficulty of the work.

Whether different levels of effort can be distinguished, and how valuable these tests may be for measurements in the field under poorly controlled conditions, remains to be discovered. Pulse rate is, of course, more easily measured than blood pressure, and it can be recorded continuously during work by means of the cardiotachometer or other devices. We can accept the statement that the pulse rate is increased by mental work in comparison to a *resting* value obtained when the subject is relaxed and calm. The changes may be small, however, and difficult to observe except when under carefully controlled conditions and with large numbers of observations. Then too, the establishment of the resting value is difficult because of the excitement which may come with the experiment. In order to con-

[5] R. D. Gillespie, The relative influence of mental and muscular work on the pulse rate and blood pressure, *Journal of Physiology,* 1923-1924, *58,* 425-432.

trol the latter it is necessary to work with subjects who are not upset by the fact that an experiment is going on.

Even with these precautions, however, it is not clear whether the pulse rate could provide us with *differential* rates for different methods of doing a task or different conditions of work. Most of the experimentation has been concerned with the question of whether or not the pulse is affected significantly by any sort of mental work, in comparison with rest or no work.

One study of visual fatigue (a special topic to be discussed more fully later) throws some light upon this question of the differential effect of difficulty of the task. Bitterman [6] found that pulse rate was reliably increased by a number checking test even with a relatively short rest period (fifteen to twenty minutes) before the work began. In reading a popular book, however, the pulse rate was decreased during reading, and fell throughout the reading period. This might be interpreted as a difference of effort, since one task involved a speed requirement while the other involved reading an interesting book at a comfortable pace.

Other results, however, failed to show that pulse rate is differentially significant. The experiment on number checking consisted of two parts, in one of which the test was in smaller type than in the other. Both tests increased the pulse rate significantly over resting values, but the difference in effect was not significant. The average pulse rates were, respectively: rest, 73.3; large type, 79.3; small type, 80.0 This was in spite of the fact that the small type was sufficiently more difficult to change the rate of work (test scores) in a significant manner.[7]

It is very likely that a useful circulatory index will involve both heart rate and blood pressure, as Lovekin has proposed.[8]

Lovekin used the combined index of *pulse-product* (pulse rate multiplied by the pulse pressure) as an index of effort. In experiments with small numbers of subjects, this index was correlated with oxygen consumption and lactic acid formation in muscular work. The relationship was considered close. In very brief experiments with other activities, the pulse pressure appeared to be correlated with emotion and attention.

---

[6] M. E. Bitterman, Heart rate and frequency of blinking as indices of visual efficiency, *Journal of Experimental Psychology,* 1945, *35,* 279-292.

[7] Pulse rate is affected significantly by noise, however. *Cf.* M. E. Bitterman and E. Soloway, The relation between frequency of blinking and effort expended in mental work, *Journal of Experimental Psychology,* 1946, *36,* 134-136.

[8] O. S. Lovekin, The quantitative measurement of human efficiency under factory conditions, *Journal of Industrial Hygiene and Toxicology,* 1930, *12,* 99-120, 153-167.

Lovekin applied the pulse-product index to workers in factory production. After allowing for the effects of meals and rests, he decided that the general trend in pulse-product was downward throughout the working day. Fatigue was indicated by decreasing production and decreasing pulse-product. An index of efficiency obtained by dividing the production rate by the pulse-product at each period of the day indicated no consistent trends in efficiency as the working day progressed. In fact, the highest efficiency occurred in the ninth, out of the ten periods into which he divided the day.

The same author gives an example of the use of this efficiency index for comparing different methods of work. In a certain task, total production was higher when there were no rest periods. He found, however, that while the production fell off 5 per cent with the rest periods, the pulse-product was reduced by 16 per cent. He concludes, therefore, that the rest periods were more efficient, in spite of the reduction in production rates.

Lovekin's report represents an interesting exploratory study, but the evidence which is offered to support the use of pulse-product as an index of effort is too scanty to permit us to regard it as anything more than an interesting suggestion.

So far, there is no convenient method for obtaining *continuous* recording of blood pressure. It must be taken periodically and is subject to temporary variations due to a variety of factors. Therefore frequent readings and large numbers of readings are essential to significant information upon the circulatory effects of work. The readings usually interrupt the work, even if they do not disturb the worker.

While there is a place for much more research into this general problem, the present situation can only lead us to conclude that there is no established technique of estimating effort by means of circulatory changes.

Muscular Tension as an Index of Effort

There is a strong possibility that measures of muscular contraction will furnish us with a sensitive and versatile index of effort. This statement can be made in spite of the objections which we have set forth in preceding pages against *identifying* alertness, attention, or problem solving with patterns of muscular contraction. We have surely not described all there is when we describe the pattern of muscular activities involved in any given activity, no matter how complete this description may be. It is quite likely that the circulatory activi-

ties are more complex than would be suggested when we think of them only as supply lines for the muscles. In addition, little is known of the central nervous system's contribution to effort, and even the digestive and glandular systems may be involved in some way.

Nevertheless, muscular tensions constitute one important aspect of total bodily functioning; they are to be regarded as an important bodily resource in psychological activity. In addition, these tensions are more amenable to experimental study than some of the other bodily activities. They have been subjected to a great deal of preliminary research within recent years.[9]

The following relationships have been studied:

1. Pressure exerted in performing the movements of a reaction test.[10] It is found that noise, for example, tends to increase the pressure exerted upon the reaction keys. This is taken to mean that additional effort is used when the work is carried on under more difficult conditions.

2. Recording of electrical potentials by amplifiers. Electrical potentials developed in various muscle groups have been recorded directly by means of special amplifiers. By means of this technique, Davis (as shown in Figure 12) has supported Morgan's conclusion concerning increased muscle tension during noise.[11] He has also shown that these patterns change with an increase in difficulty of the task.[12] Potentials from both arm and neck increase with the difficulty of the task. (Those problems which were failed by a larger proportion of subjects were considered the more difficult.)

The pattern of distribution of the tensions in various parts of the body varies with the kind of activity in which the subject is engaged.[13] During mental arithmetic the greatest amount of muscle activity was found in the right arm. The other two regions studied (the other arm and one leg) also showed increased activity, but not so much. In learning nonsense syllables, the activities of the three regions were more nearly equal, the *left* arm showing the greatest contraction. The

---

[9] See F. A. Courts, Relations between muscular tension and performance, *Psychological Bulletin,* 1942, *39,* 347-367, for a summary of this field. See R. C. Davis, Methods of measuring muscular tension, *ibid.,* 329-346, for a survey of methods of measuring muscular tension.

[10] J. J. B. Morgan, The overcoming of distraction and other resistances, *Archives of Psychology,* 1916, No. 35, p. 84.

[11] R. C. Davis, The relation of certain muscle action potentials to "mental work," *Indiana University Science Series,* No. 5, 1937.

[12] R. C. Davis, The relation of muscle action potentials to difficulty and frustration, *Journal of Experimental Psychology,* 1938, *23,* 141-158.

[13] R. C. Davis, Patterns of muscular activity during mental work and their constancy, *Journal of Experimental Psychology,* 1939, *24,* 451-465.

control of the various muscle groups is not, however, independent. Davis reports that subjects instructed to relax the right arm during mental multiplication or learning relaxed the muscles of other parts of the body as well. When the relaxation occurred, the output seemed to be reduced.

While the direct measurement of potentials from various muscle groups is the most clearly valid method of measuring muscular tonus or tension, it has certain disadvantages. The method is confined to

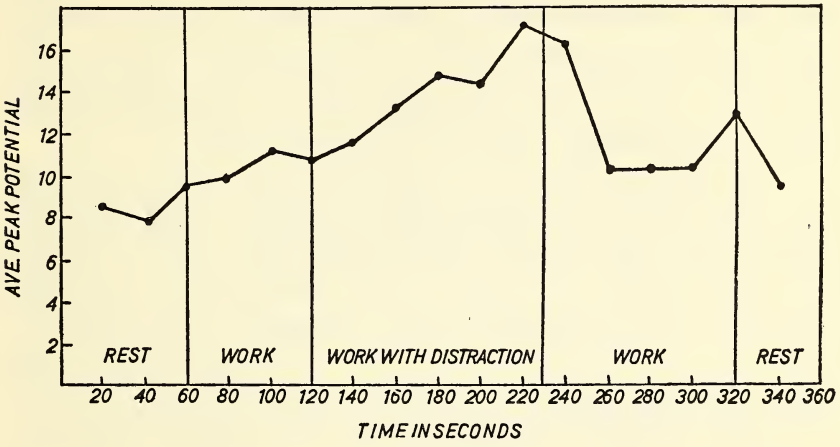


Figure 12. Action Potentials During Rest and Work (Reading), With and Without Distraction

(From Davis, *op. cit.*, 1937, p. 24.)

laboratory applications because of the amount of apparatus involved, and because the subjects must be electrically shielded to prevent potentials from other sources entering the amplifiers and confusing or covering up the records.

Even in the laboratory, the method is subject to some disadvantages. The muscle potentials involved in complicated activities may be all over the body. The activity of one muscle group cannot be taken as representative of what is happening in other muscles. The focus of activity may shift from one part of the body to another with changes in conditions of work. One muscle group might, therefore, show a high level of activity while the average tension of the whole body is low.

To measure *total* effort as reflected in muscle tonus it would be necessary to record simultaneously the potentials from a number of different parts of the body. This would be a difficult and expensive process. Perhaps it will be discovered that there are certain muscle

groups which can be taken as representative of larger areas of the body, and thus it will be possible to reduce the number of simultaneous observations required. Much research remains to be done, however, before we reach this point in the application of amplifier techniques to the study of efficiency.

Several studies of practical problems by this method have been inadequate because they involved only one or two muscles. It has been found, for example, that work during loud noise at first raises the potentials of the forearm, and that then the potentials gradually decrease, but we have no way of telling whether or not the tensions have moved to some other part of the body, or discovering that there has been a bona fide reduction in over-all tension. This criticism would apply equally well to the methods which record pressure exerted by the working member which we described previously.

3. Mechanical recording of muscle contractions. The thickening of a muscle can be recorded by means of a lever or other device attached to the body surface, or the tautness of a tendon measured by pressure applied. By these methods results comparable to the findings of electrical studies have been reported.[14] This method requires less expensive apparatus, and the subjects do not, of course, have to be electrically shielded. The apparatus is likely to restrict the movements of the subject and sometimes it has required the subject to maintain a somewhat artificial position. In addition, these mechanical methods measure the activities of a single muscle group, and so are subject to the same disadvantages as the electrical methods described in the preceding paragraphs.

4. Electrical resistance of the skin. The electrical resistance of the skin, as measured by ordinary physical methods, shows variation with the activity of the subject. There is now evidence that as the subject becomes more active muscularly, as he becomes involved in an emotion, or as he engages in mental work, the electrical skin resistance is reduced. Consequently, skin resistance is possibly an index of effort—the lower the resistance, the greater the effort. Let us examine the concrete evidence upon which this suggestion is based.

It has long been known that the electrical resistance of the skin falls rapidly after sudden stimuli (loud sounds, electric shocks) or when embarrassing questions are asked. In fact, these sudden changes in resistance are used as the basis of one of the common methods of lie detection. These fluctuations are now generally re-

---

[14] G. L. Freeman, The spread of neuromuscular activity during mental work, *Journal of General Psychology*, 1931, *5*, 479-494.

garded as a function of sweat gland activity which is controlled by the sympathetic nervous system—a portion of the nervous system which has much to do also with the control of circulation, digestion, and other "vegetative" functions.[15] The palms of the hands and the soles of the feet seem to be more easily influenced than other portions of the body surface by slight changes in the activity of the subject, although the changes do occur in other surfaces also.[16]

As we have indicated, the earlier studies of electrical skin resistance were concerned with short-term, relatively sudden changes in resistance. This phenomenon has become known as the "psychogalvanic response" (PGR). More recently, however, it has been observed that the *general level* of resistance varies more gradually, and that this general level also has significance.

The level of skin resistance is now often regarded as dependent upon the general level of excitation of the muscles of the body. It is therefore hoped that it can provide us with an index of the total muscular activity and tension of the body as a supplement to the information about specific muscle groups which is given us by other methods.

This hope is based upon a number of lines of evidence which we may examine briefly:

(*a*) The resistance changes with gross changes in muscular effort or relaxation. For example, the skin resistance is definitely lower when

---

[15] R. S. Woodworth, *Experimental Psychology,* New York, Henry Holt & Co., 1938, pp. 276-297.

[16] C. W. Darrow and G. L. Freeman, Palmar skin resistance changes contrasted with non-palmar changes and rate of insensible weight loss, *Journal of Experimental Psychology,* 1934, *17,* 739-748. In this article the authors conclude that palmar resistance changes are correlated with changes in alertness, anticipation, or apprehension, while non-palmar changes are correlated with the amount of muscular activity. That is, the non-palmar sweat glands are mechanisms of heat dissipation, while the palmar glands have a different function. In a later article, however, Freeman and R. M. Simpson (The effect of experimentally induced muscular tension upon palmar skin resistance, *Journal of General Psychology,* 1938, *18,* 319-326) apparently contradict this conclusion, stating that "the different skin areas represent largely differences in the ease of excitation, with the palm becoming more active first and dropping toward its limit of physiological reactivity as the non-palmar areas become increasingly responsive." This is in spite of the definite statement of Darrow and Freeman in the previous article that they did *not* find that the palmar resistance was always low when the non-palmar areas became active.

This question has an important bearing upon the question of what the skin resistance indicates, and the role of muscle tension in psychological activity. If we identify psychological activity and muscular activity, then the only question is, how best to get an index of activity of the muscles from the skin resistance. If, however, we grant that there are other bodily systems involved independently in any psychological activity, it might be very important to find out whether the palmar resistance indicates a kind of internal activity different from that related to non-palmar changes. The palmar changes might, for example, be more closely related to circulatory and digestive functions than to skeletal muscular activity.

the subject is sitting up than it is when he is reclining.[17] It is not known whether the resistance is proportional to the amount of relaxation or only roughly correlated. Wenger and Irwin, for example, found that the resistance was not much lower in an extreme state of tension than it was when the subject was sitting up. Darrow finds that a relatively slight change at low resistance levels may indicate as much physiological change as a large variation when the resistance is high.[18] He therefore proposed the use of the logarithm of the *conductance* $(1/R)$, believing that this measure is more nearly proportional to the tension than resistance (R) as measured in ohms..

It may be that it will be necessary to use the palmar resistance for some part of the total range of tension, while using non-palmar resistance to indicate changes in the more extreme parts of the range.[19]

(*b*) The skin resistance falls during sedentary activities of various kinds (mental arithmetic, solving problems from an intelligence test, etc.). There seems to be some evidence that the change in electrical resistance is greater if the problem is more difficult, or if the subject is "frustrated" in some way. The latter evidence is, however, still incomplete, and much study is needed in order to determine the relationship of difficulty of the task, or the effort required, and the level of resistance. Freeman, for example, compared the resistances while the subject worked at a "comfortable pace" which the subject had chosen, with the resistances during work at a slower or faster pace.[20] While all the differences were in the expected direction, none of them were fully significant statistically. In another experiment involving sensory discriminations, it was possible to measure the difficulty of the discriminations in terms of the number of failures in a certain comparison.[21] Here there was a clear-cut correlation between level of difficulty and lowered resistance, but no statistical analysis is reported.

Another experiment of Freeman's illustrates another phase of this relationship.[22] In an extended series of reaction time measurements

---

[17] M. A. Wenger and O. C. Irwin, Fluctuations in skin resistance of infants and adults, and their relation to muscular processes, *Iowa University Studies in Child Welfare,* 1936, *12,* No. 1, 143-179.

[18] C. W. Darrow, The equation of the galvanic skin reflex: I. The dynamics of reaction in relation to excitation background, *Journal of General Psychology,* 1937, *16,* 285-309.

[19] See footnote 16, p. 107.

[20] G. L. Freeman, The relationship between performance level and bodily activity level, *Journal of Experimental Psychology,* 1940, *26,* 602-608.

[21] G. L. Freeman and W. J. Giese, The relationship between task difficulty and palmar skin resistance, *Journal of General Psychology,* 1940, *23,* 217-220. The result here may be a special one, since the discriminations were given in ascending order of difficulty within the same period.

[22] G. L. Freeman, *op. cit.,* 1940.

with the same subject, it was found that the fastest average reaction time accompanied a moderately low resistance. If the resistance dropped still lower, the reactions were irregular and the average time was slower. The reactions were also slow at those times when the resistance was high. Thus we cannot determine the relationship of resistance to effort by measuring output. The individual may be working hard, perhaps too hard, and still show a poor performance. The difficulty of the task, or the effort required, must be measured in some manner independent of the output of any given subject.

In addition to these lines of evidence upon specific methods of measuring tension, there have been a number of researches concerned with the effect of tension upon performance. It has been found, for example, that rate of learning and performance on some other tasks can be influenced by artificially adding to the total muscular tension of the subject.[23] If, while working, the subject is asked to compress a spring or support a weight with one limb, his performance may be increased unless the additional tension becomes too great.

Thus it appears that even extraneous tension which might appear to be "waste effort" may enhance performance. Just what the mechanisms are is a subject of considerable discussion which need not concern us here.

Bills takes the position that efficiency depends upon the general tension level of the body.[24] If this level is either too low or too high, efficiency goes down. It is certainly indicated by these experiments that efficiency can be impaired by levels of tension which are too high. Here, more effort is being expended while the output is reduced. When output is reduced by too low a level of tension, the net influence upon efficiency is less clear. The net effect depends upon the relative loss in output and the amount of reduction in tension. Thus we may say that a person under induced tension learns *faster* than he does under normal tension, but to say that he is more efficient goes beyond the information available. There is a good deal of confusion in discussions of learning because in the context of learning studies *time required to learn* has often been taken as the sole element of the cost of work. This is, of course, merely equating efficiency and rate of output.

---

[23] F. A. Courts, Relations between muscular tension and performance, *Psychological Bulletin,* 1942, *39,* 347-367.

[24] Bills, *op. cit.,* pp. 22-25. The conclusion that too little tension also decreases efficiency may be a confusion of the terms *efficiency* and *output* in Bills' discussion of this point. (Bills does not include "efficiency" in his glossary of terms or in his index, and seems to vary somewhat in his use of the term.)

Summary.—From the evidence now available, it appears that measurements of muscular tension furnish us with our most promising line of approach to the measurement of effort in those tasks which are commonly classified as light manual work and mental work. Electrical potentials from the muscles can be measured in the laboratory, where special foci of muscle contraction are of interest. It may be that some form of composite measurement will permit us to evaluate the total muscular activity during work, just as we can get a measure of the average temperature of the whole surface of the body.

There is also considerable promise in electrical skin resistance as an index of the over-all activity level, with the resistance going down as the activity increases. Whether it indicates muscular tensions alone, or whether it reflects the general level of internal activity of all kinds, it is probably a useful index, since all these activities are to be reckoned in the cost of work. In fact the index might be even more useful if it reflects more than the level of muscle tension alone.

Skin resistance can probably be used immediately if it is coupled with other evidence as a check. There is still much fundamental research to be done in determining exactly what it measures, how much influence relatively unimportant factors may have upon the level of resistance, and how accurate it is as a measure of *amount* of effort as well as a reflection of relative changes of effort.

Some Additional Physiological Changes.—Heavy muscular work produces many changes in physiological functions besides those which we have touched upon in this book. It is natural that some of these changes should be studied in connection with sedentary work as well. As an example, we may mention briefly the studies of blood count as affected by work, since some significant effects of sedentary work have been claimed in this aspect of function.

Muscular work produces increases in the numbers of both red and white blood cells, with some evidence of a relative as well as an absolute increase in the number of white cells (leucocytosis).[25] Moss and his collaborators found that fifteen minutes of activity on the bicycle ergometer increased the white cell count by more than 50 per cent.[26] It has been claimed that "mental" work also produces leucocytosis, perhaps even more marked than that observed in muscular work.[27] Laboratory studies involving psychological test activity for

[25] Atzler, *op. cit.*, 1938, 372-375.

[26] F. A. Moss, J. H. Roe, O. B. Hunter, L. French, and T. Hunt, The measurement of fatigue by physiological methods, *Journal of Experimental Psychology*, 1931, *14*, 423-438.

[27] R. M. Skliankaia, C. E. Shakhnovich, A. E. Levina, and A. I. Zak, Changes in the composition of blood during mental and light physical work, *Journal of*

two hours, as well as occupational studies of telephone operators and hosiery knitters, were the basis of these conclusions. Even in these positive results, however, there are irregularities which signal caution in interpretation. The hosiery knitters, for example, failed to show the increase in white cells on the night shift, and the authors point out the great irregularity of changes from individual to individual.[28] There are, perhaps, diurnal variations as well as other unpredictable factors to be taken into account. In addition, there is little correlation with the duration or amount of work performed.

In a study of truck drivers who had been on the road for varying numbers of hours, the white count for those who had been driving was compared with the count for those who had not been driving.[29] While there was a greater number of white cells for those who had been driving as compared with those who had not driven at all, there was no clear change in the white count as the number of hours of driving increased. In some groups, in fact, the group who had been driving for ten hours showed a lower count than those who had been driving for shorter periods.

Besides these irregularities of the effect of sedentary work upon the blood count, irregularities which make the count a doubtful measure of effort or of fatigue, there have also been researches with completely negative findings. Dunajewski and Kaplan could find no effect of three hours of work at various sedentary tasks upon ten subjects.[30] At least there was no effect which was greater than the normal "random" variations in the blood structure.

Since changes in internal bodily economy are also reflected in the chemical composition of the urine, a number of tests involving urine-analysis have been employed. Until recently, however, there appeared to be insufficient regularity in these findings to provide a quantitative indicator of the cost of work, especially in the more sedentary types of occupation. Pincus and Hoagland have reported promising results with a new type of test, based upon analysis of certain urinary com-

---

Industrial Hygiene and Toxicology, 1938, *20,* 169-178. G. A. Lewitina, A. J. Lewina, O. S. Tschernomordik, K. S. Samytschkina, L. M. Sidorowa, and S. S. Shapiro, Der Einfluss geistiger Arbeit auf das weisse Blutbild. Ein Beitrag zur Frage des neurogenen Ursprungs der Verschiebung in der Schillingschen Formel, *Arbeitsphysiologie,* 1931-1932, *5,* 115-124.

[28] Skliankaia, *et al., op cit.*

[29] B. F. Jones, R. H. Flinn, and E. C. Hammond, Fatigue and hours of service of interstate truck drivers, U. S. Public Health Service, *Public Health Bulletin No. 265,* 1941, 53-57, 227-240.

[30] M. J. Dunajewski and P. M. Kaplan, Über das Blutbild bei geistiger Arbeit, *Arbeitsphysiologie,* 1932-1933, *6,* 437-444.

ponents derived from glandular secretions.[31] These products derive from the adrenal cortex and are called 17-ketosteroids. Products of this gland have long been known to relate to the processes of muscular activity, fatigue, and metabolism.

The amount of steroid excreted in each sample is compared with the amount excreted at the same time of day on a control or non-working day, since the excretion varies from a high rate in the morning to a low rate at night. In studying the effect of work, this normal variation is allowed for, and the excess is correlated with the activities of the experimental subject.[32] The experimenters found that a prolonged activity upon a test which simulated the activities of flying produced a marked excess of ketosteroid excretion toward the end of the period. This excess was correlated with a parallel increase in errors in the test. Later, similar results were observed in pilot instructors and test pilots actually engaged in flight. In the flying instructors, the correlation between excess ketosteroid and per cent of the time spent in the air was high ($r = .978$). The individual pilot's rated ability to "take it" was also correlated with his characteristic individual rate of ketosteroid excretion ($r = .676$). It was also noted that the total urinary output was correlated with the amount of flying time or time at work in the flying test. The relationship of this diuresis to errors or to flying time was not so close, however, as that observed for the ketosteroid excess. As a further check upon the validity of their findings, Pincus and Hoagland have performed controlled studies of the effect of administering a synthetic steroid to working subjects.[33] By this means they were able to reduce the errors in the flying test which normally occurred toward the end to the test period. They have also found significant increases in factory productivity resulting from the administration of this substance.[34]

It would appear that this discovery will be valuable for the study of occupations involving strong emotional stress coupled with exacting, controlled muscular activity. That is as far as we are justified in going at the present time. To show that the test is useful for indus-

---

[31] G. Pincus and H. Hoagland, Steroid excretion and the stress of flying, *Journal of Aviation Medicine,* 1943, *14,* 173-193. H. Hoagland, Adventures in biological engineering, *Science,* 1944, *100* No. 2587, 63-67.

[32] The "normal" curve is based on results for seven subjects in ordinary daily activities (uncontrolled); Pincus and Hoagland, *ibid.*

[33] G. Pincus and H. Hoagland, Effects of administered pregnenolone on fatiguing psychomotor performance, *Journal of Aviation Medicine,* 1944, *15,* 98-115, 135.

[34] G. Pincus and H. Hoagland, Effects on industrial production of the administration of Δ 5 pregnenolone to factory workers: I, *Psychosomatic Medicine,* 1945, *7,* 342-346. G. Pincus, H. Hoagland, C. H. Wilson, and N. J. Fay, Effects on industrial production of the administration of Δ 5 pregnenolone to factory workers: II, *ibid.,* 347-352.

trial tasks, we should need answers to the following questions: (1) Do similar excretions occur as regularly in types of work in which the stress is due to distaste for the work, awkward postures, and static strains, "intellectual" work or problem solving, and so on. The results of administering steroids do not indicate the sensitivity of steroid output as an index of fatigue. (2) Are the individual differences in excretion so great that we must use very large groups of subjects in a study of methods of work, or the effects of certain factors upon work? (3) How representative is the "normal" curve which these investigators use in computing all excess values? Variations in this "normal" curve might produce a spurious appearance of excess excretion toward the end of the day. This could not account for the results when excretion was correlated with percentage of time spent in flying, unless there were a tendency for more flight in the later hours. In other studies, however, excess is correlated with time spent in work. If there is at the same time a progressive error in computing excesses, it would produce a spurious correlation. (4) Is the indicator sensitive enough to reflect variations in a single factor, or will it reflect only changes in several conditions at once? Pincus and Hoagland have shown that the steroid excess is increased when the work is carried out with a serious deficiency of oxygen. Would it be similarly affected by changes in temperature, lighting, noise, and the like, and how sensitive would it be to variations in these factors? (5) How cumbersome and expensive is the method? [35]

These authors have mentioned that they were making some industrial studies which may help to answer some of these questions. At the present writing these results have not yet become available.

Another measure, which falls somewhere on the border line between performance tests and physiological measures, is the critical flicker-fusion frequency. This is measured by asking the subject to watch an area which is illuminated intermittently. The frequency of the flashes of light is gradually increased until the subject reports that he no longer sees any flicker. The test may also be conducted with varying *brightness* of the flickering light as well as with varying *frequency,* since the perception of flicker is a function of both variables. As the brightness increases the critical frequency necessary to eliminate flicker also increases. (See Figure 13.)

---

[35] This fifth question is not, of course, of equal importance with the other four. If a highly reliable and valid technique can be discovered, the gains from its use would ultimately justify its expense if no other method were available. Nevertheless, the speed with which advances can be made will depend very definitely upon this factor.

The test originated as a device for studying certain aspects of the visual mechanism, but it has recently been brought into the context of fatigue. Interested in the possible fatiguing effects of motion pictures upon the ocular mechanism, Snell measured the critical frequency before and after a period of exposure to a flickering

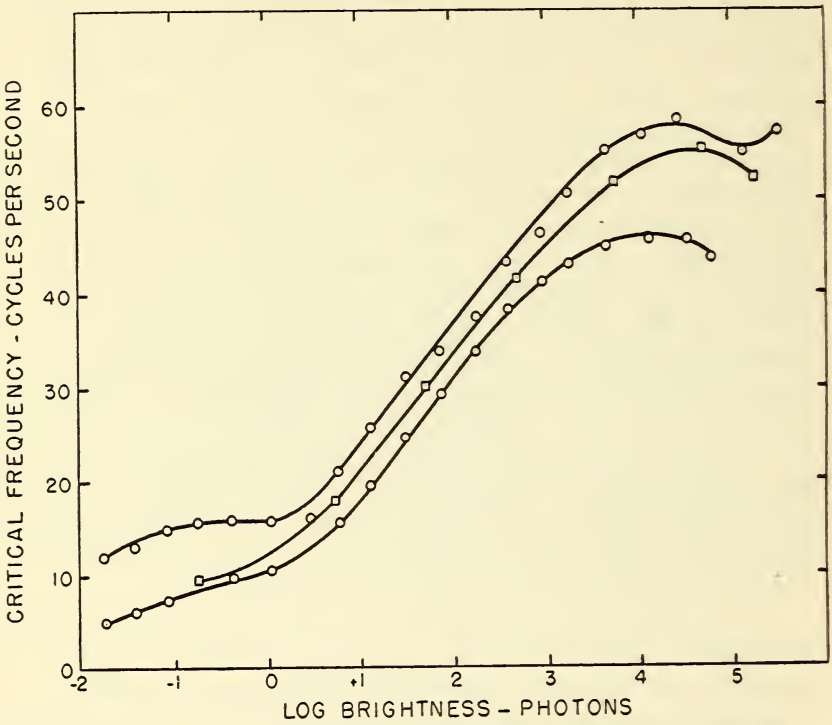


Figure 13. The Relation between Critical Flicker-Fusion Frequency and Brightness

Lower and upper curves for areas of 6° and 19°, respectively, from data of Hecht and Smith. Middle curve for area of 12° for one of Lee's subjects.

(From R. H. Lee, Critical fusion frequency of flicker, U. S. Public Health Service, *Public Health Bulletin No. 265*, 1941, p. 195.)

stimulus.[36] He found that exposure to visible flicker had the effect of reducing the critical frequency; that is, a rate of alternation which would be seen as flickering before the exposure would later appear to be continuous. According to Snell, it made no difference what rate of flicker produced the "fatigue," so long as it was visibly flickering. This finding is supported in general by McFarland and his collaborators, although they varied the intensity of the test stimulus rather

[36] P. A. Snell, An introduction to the study of visual fatigue, *Journal of the Society of Motion Picture Engineers*, 1933, *20*, 367-390.

than its frequency.[37] The latter investigators found that a slowly flickering stimulus produced a greater effect in a given time interval than a more rapidly flickering light.

More recently, Simonson and Enzer have proposed the use of the flicker fusion test as a measure of "general" fatigue of the central nervous system, removing the test from the context of specific ocular fatigue.[38] In a number of office, laboratory, and professional workers they found a decrease of the flicker fusion frequency in the afternoon as compared with the morning value, the decrease occurring in every case they studied. They also found that benzedrine administered to the subjects during the day always eliminated the effect of the day's work upon the fusion frequency.[39]

Simonson and Enzer's findings were so uniform that they gave great promise to the technique of flicker fusion testing. Unfortunately, however, their results were not duplicated in a similar experiment carried out by McFarland and his collaborators.[40] Seven of the nine subjects used by McFarland were engaged in reading examinations for the whole working day, and the other two were employed in computing and drawing. In most cases there was no difference in the results of the flicker test in the afternoon as compared with the morning. Those differences which did occur were in opposite directions with about equal frequency.

There was only one result in McFarland's study which would support Simonson's contention that the flicker fusion test is a measure of fatigue of the central nervous system (or of any sort of fatigue in sedentary work). Subjects who had gone without sleep for a period of at least twenty-five hours showed some individually significant changes in the frequency-intensity relationships of the flicker test. (The five tests upon three subjects were divided as follows: One subject, tested once, showed no change following the sleepless period. A second subject, tested twice, showed a significant change on one test but not on the other. The third subject showed significant changes on both occasions.) On the basis of these rather meager results, McFarland concludes that "loss of sleep appears to produce a significant change in the flicker curves. Following loss of sleep, a

---

[37] R. A. McFarland, A. H. Holway, and L. M. Hurvich, *Studies in Visual Fatigue,* Harvard University: Graduate School of Business Administration, 1942, pp. 191-219.

[38] E. Simonson and N. Enzer, Measurement of fusion frequency of flicker as a test of fatigue of the central nervous system, *Journal of Industrial Hygiene and Toxicology,* 1941, *23,* 83-89.

[39] E. Simonson, N. Enzer, and S. S. Blankenstein, Effect of amphetamine (benzedrine) on fatigue of the central nervous system, *War Medicine,* 1941, *1,* 690-695.

[40] R. A. McFarland, et al., *op. cit.,* 196-204.

greater intensity is necessary at a fixed frequency for flicker recognition to occur throughout most of the range used. There are, however, considerable individual differences." [41]

It is possible that some of the differences in results of Simonson and of McFarland are attributable to differences in technique. Simonson varied the frequency of the flashes at certain fixed intensities of the light, and determined the threshold of recognition of flicker in terms of frequency. McFarland varied the intensity at certain fixed frequencies to determine thresholds in terms of intensity. A curve like that shown in Figure 13 could, of course, be plotted from either kind of data, since it would not matter which variable is considered dependent. The difference in technique may, however, make a real difference to the subjects, and it is therefore still possible that fatigue would show its effects upon one type of threshold and not upon the other.[42] It is unfortunate, therefore, that the tests have not been carried out by both methods upon the same subjects and under the same working conditions.

Using a method more closely similar to that of Simonson, Lee measured the fusion frequencies of truck drivers who had been driving for varying periods before the test.[43] (This study was a part of the program mentioned previously in connection with blood tests for fatigue (page 110); it is discussed again in a later chapter on accident problems (Chapter 12). There was a slight decrease in the critical frequency as the number of hours of driving increased, but none of the results were statistically significant. Again it is difficult to interpret the results because there was no control test before driving for the specific driver. One group of drivers had not driven at all, while *another* group had driven before the test. The failure to find significant differences may, therefore, be due to the increased variability of the data resulting from individual differences. It is notable that the differences between drivers in different *cities* were almost as large and as consistent as the differences between drivers grouped according to the number of hours of previous work. This suggests other important variables which have not yet been taken into account, and which must be discovered before the results of such tests would be meaningful.

---

[41] *Op. cit.*, p. 217.

[42] That is, even if the two methods give equivalent results for unfatigued subjects, fatigue might have different effects upon the two kinds of judgment.

[43] R. H. Lee, Critical fusion frequency of flicker, in B. F. Jones, R. H. Flinn, *et al.*, Fatigue and hours of service of interstate truck drivers, U. S. Public Health Service, *Public Health Bulletin No. 265,* 195-208.

An entirely different application of the flicker-fusion test was tried out by Brozek and Keys.[44] Since Simonson had supported the theory that flicker is affected by general fatigue of the central nervous system, Brozek and Keys measured fusion frequencies for subjects who were active in heavy muscular tasks. They first measured the reliability of the test by comparing repeated measurements for the same subjects, and found the test to be highly consistent and unaffected by practice. Heavy muscular work, however, had no effect upon the fusion frequency except under very extreme conditions. The frequency decreased slightly from day to day when the subject was engaged in work nearly at capacity in a temperature of 120° F. Even then, however, there was no significant change from morning to night of the same day. There was a similar tendency for a long-term decrease for subjects who were working hard with inadequate food intake. None of the changes observed in these studies were large, and frequently they were not statistically significant.

Another study indicates that flicker fusion is related in some fashion to general bodily economy. In analyzing the effects of oxygen deficiency in altitude chambers which simulate the conditions of flight at varying altitudes, it was found that flicker fusion frequency was reduced.[45] The changes were significant at an "altitude" of 10,000 feet. For higher altitudes, there was also a significant residual effect, in that the frequency remained low for a time after returning to "sea level."

We may conclude this brief summary of the flicker-fusion test by stating that, in spite of the conflicting results now available, it is a subject worthy of further serious and well-controlled research. Perhaps the test will be useful only in studies of certain visual tasks. Perhaps Simonson's belief in its value as a measure of fatigue of the central nervous system will be further supported by other kinds of experiments. At any rate, the test will be worth the effort which may be devoted to it, since it possesses important advantages over many performance tests. The subject has no way of judging the accuracy of his own performance—in fact there are no "correct" or "incorrect" reactions, and even if there were the subject cannot guess the way in which the apparatus is manipulated by the experimenter. It is likely, also, that the motivation of the subject would have less effect

---

[44] J. Brozek and A. Keys, Flicker fusion frequency as a test of fatigue, *Journal of Industrial Hygiene and Toxicology,* 1944, *26,* 169-174.

[45] J. E. Birren, M. B. Fisher, E. Vollmer, and B. G. King, Effects of anoxia on performance at several simulated altitudes, *Journal of Experimental Psychology,* 1946, *36,* 35-49.

upon these results than it would for most performance tests. It is primarily for this reason that we have discussed the test in the context of physiological measures.

Gross Statistical Measures of Efficiency

The measures of cost of work and of efficiency which we have considered so far are expensive and difficult to apply. They look toward future solutions of practical problems by analyzing the effect of work upon bodily processes of all kinds. Where more immediate answers to practical questions of efficiency have been desired, students of industrial problems have often resorted to cruder measures which are more easily applied. Thus the effects of rest periods, various lengths of the working day, changes in lighting or ventilation methods, and of supervisory practices have been studied more frequently without utilizing any of the methods so far discussed, except, possibly, for an analysis of daily work-curve patterns.

Several of the measures listed below would appear to confuse efficiency with rate of performance. In fact, that confusion has often been made in industrial studies. In other cases, however, the interest has centered in efficiency in the more correct sense of the term, and these measures can be justified as inexact indices of efficiency under certain conditions. There are so many possible measures that we shall merely list them briefly, and then consider the arguments which may be offered to justify the more important ones:

1. Long-term trends in production rate.
2. Records of errors, waste, scrap, etc.
3. Accident rates.
4. Turnover rates.
5. Records of health and absenteeism.

For all these indices, a prime requisite is *adequate* statistical treatment beyond the mere collection of numbers and averages. This seems self-evident, but in practice the requirement is often overlooked. There are instances on record in which the optimal arrangement of rest periods was determined from a small group of subjects and without any analysis of the variability of these subjects or their representativeness. In other words, it would not be possible to conclude that the results are significant *for that group,* to say nothing of the application of the results to other individuals doing the same kind of work.

The reader may feel that we harp upon this theme to an excessive degree. It is apparently necessary, however, to emphasize that statistical treatment to determine the significance of results in relation to chance is always essential.

With this general condition, let us examine some of these indices in greater detail in order to note the assumptions upon which their use is based, and to evaluate these assumptions.

1. Long-Term Trends in Production Rate. Assuming that there has been no change in motivation, it is argued that production must be increased as a result of lowered cost per unit. If factors other than learning are under study, it would also be necessary to ensure that the influence of practice has been eliminated. Ordinarily, this is not taken into account, apparently because it is believed that the workers have already reached the limits of improvement due to practice.

The question here is, what is meant by "constant motivation." Usually it is assumed that it is the same if the pay remains the same, and if there is no evidence of marked changes in supervisory methods, etc. It is assumed that under these conditions the worker will expend a fixed amount of energy per unit of time. If the rate of production increases, it must mean an increase in efficiency.

There are two possible flaws in this reasoning. First, the factor under study may itself provide a change in motivation so that the men are working harder after this factor has been introduced. There is no way of determining whether or not this is the case without a more direct measure of effort expended. Secondly it might be equally reasonable to assume that constant motivation implies that the worker would aim at a given amount of production or a given amount of money in the pay envelope. In other words, the factor under study may have more effect upon effort expended than it does upon the gross output. If so, the factor would appear to have no effect if our only measure is the output trend.

There are two general arguments which might be used as a retort when we mention the above objections to studying efficiency by analyzing production trends.

(*a*) The attitude of industrial managers may be that they are not interested in these problems. If production goes up, they do not care whether it is due to a motivational effect or to an increase in efficiency. They would also argue that an increase in effort cannot be seriously harmful if the increase in output is maintained over a long period of time. A temporary increase followed by a slump would, however, indi-

cate that the method or factor involved was so inefficient as to be valueless.

(*b*) If, on the other hand, production remains the same, management might not feel that it could justify any expense involved in using the new factor or method, even if the worker's effort was reduced.

The first of the above points is probably justified, so long as the production index is followed over a long period of months or years. This would mean, however, that the determination of the effects of a given factor would take an extremely long time. The second point, that management is not interested in reducing the effort unless it is reflected in output is not so easily justified. If the effort required for a unit of work can be reduced, it should be to management's advantage to do so. For one thing, it would probably pay in terms of reduced overhead or indirect costs such as turnover, complaints, discontent, and the like. Another reason why it might pay is that it provides an additional reservoir of productive capacity which could be tapped if suitable incentives were provided along with the changed conditions of work.

We may conclude, therefore, that if a given factor increases production consistently over a long period of time, efficiency has probably been increased; or, at least, that management would be justified in using that factor. If, on the other hand, no increase is observed, no clear conclusion is possible. Production trends, then, are useful as positive evidence, but not as negative evidence unless other indices are also studied.

2. Records of Errors, Waste, Scrap, Etc. If production rates are controlled by automatic machinery, or are otherwise insensitive to changed conditions, the effects of the changes may still be evidenced in measures of the quality of work. We have already indicated that such measures of accuracy and control are valuable in laboratory studies of production rates. Even if production rates are significantly affected, records of quality should be examined to make certain that the changed performance is a bona fide change in useful output.

3. Accident Rates. Although this is usually considered as a separate problem, accidents nevertheless represent an evidence of inefficiency. In occupations involving frequent or serious accidents, they may represent the largest part of the total cost of the work, and energy cost would be a relatively minor factor.

In certain occupations, then, accident rate can provide a useful indication of efficiency. There can be no doubt of its validity if records are carefully and accurately kept. A long-term collection of data is necessary, however, in order to ensure representative data. In many occupations, the accident rate would be so low that it would be almost impossible

for practical reasons to obtain sufficient data to measure the effect of any one factor. The longer the span of time involved in a set of data, the more chance there is that other variables have changed during the period of study, and the more difficult becomes the evaluation of the findings for this reason. Accident rate can be considered only as a special index to be applied to a limited number of occupations.

4. Turnover Rates. Here again we meet the difficulty that observations must be carried on over a long span of time if they are to be significant. In addition, turnover rates are subject to even more variations, due to extraneous factors which are not the immediate subject of investigation. Is it often difficult to separate motivational factors from fatigue and other elements involving capacity for work.

From the point of view of management, turnover rates evidently reflect the over-all efficiency of the organization's supervision, recruitment, or other personnel policies. Our more immediate concern is with the efficiency of performance of the worker himself. It is probable that an increase in turnover rate would constitute a gross indication that there is some increase in discontent, fatigue, or energy expenditure. It would indicate decreased efficiency, although without other information it would be impossible to determine what elements of the cost of work are primarily responsible for the change.

5. Records of Health and Absenteeism. The incidence of illness or of some particular disease is even less sensitive to variations in some factor in the work than other measures mentioned above. Of course, specific immunological treatment should show its effect upon large groups in a reasonable period of time. But if it is desired to find the effect of such a factor as overtime or long working days upon the incidence of illness, a much more extensive sample would have to be studied, and over a much longer period of time. If absenteeism is to be considered as a reflection of effects upon health, then some form of investigation of each case to determine the cause of absence would be necessary to make the data meaningful.

Each of these gross measures of efficiency is useful, therefore, under certain conditions, but only limited conclusions can be drawn. To get the maximum use from them we hope that they can eventually be studied in parallel with more exact analyses of the effect of the work upon the worker. The gross measures will continue to be useful on this basis, even after more accurate methods are available, because they will furnish us with an economical method of preliminary analysis of the problem at hand.

Summary and Conclusions on Measuring the Cost of Work

The reader has probably gained the impression that these chapters on the fundamental problems of efficiency have raised more questions than they have answered. If so, it is all to the good. One cannot have an adequate grasp of this field until he is aware of the problems which need to be solved in paving the way for more adequate solutions of practical problems.

Applied psychology is concerned primarily with practical problems, but its aim is to solve these problems by the methods and results of scientific investigation. We cannot give satisfactory answers to many practical problems of efficiency until the methods of finding these answers are developed and established. The methods which have been critically surveyed in these chapters provide a beginning from which tentative answers to some of these questions can be given.

In succeeding chapters we shall consider specific factors which affect productivity and efficiency of human performance. We shall be able to deal with many of these very briefly because we have already become familiar with the technical problems involved in measuring the effect of a given factor.

Among the measures of the cost of work which we have considered there have been three main kinds: (1) Those which show promise of developing into valid indices of fatigue or other phases of the cost of work, but which are still in the developmental stage. (2) A few measures which have more established positions as measures of the cost of work, although their applicability is limited to certain definite kinds of work. (3) Crude indices subject to a variety of objections, but which are relatively easy to apply, and are used as stopgaps, providing us with rough solutions until more refined methods become practically useful.

The first of these classes would include: muscle potentials, skin resistance as an indicator of tension, steroid excretion, flicker fusion frequency, composite measures of circulatory activity, such as the pulse-product, and possibly certain psychological test activities, especially those which require delicate co-ordinations, and closely controlled attention. The second class of more firmly established techniques would include, as its principal member, the use of oxygen consumption for heavy muscular work, possibly also for moderately heavy industrial tasks, but not for sedentary tasks.

The last class of cruder techniques will, unfortunately, provide us with the bulk of the information which is available on several of the

problems of practical import in industrial work. This class includes output decrement curves (for heavy physical labor, or other work with maximal effort), errors, variability and accidents during performance, long-term trends in productivity, and, upon occasion, statistics of illness, complaints, or turnover.

Chapter 6

SPECIFIC FACTORS GOVERNING EFFICIENCY AND PRODUCTIVITY

In the preceding chapters we have discussed some of the fundamental problems involved in evaluating the efficiency of the organism under any given conditions. We shall now survey some of the available knowledge concerning factors which affect the rate of work and efficiency of performance. In the present chapter we shall discuss conditions which serve as *governing* factors in performance, that is, they increase or decrease the capacity of the individual to perform, or they affect primarily the ease or difficulty of the "work itself" rather than controlling directly the effort which the individual puts forth in his activities.

It is true that some of these may, on occasion, play a dual role, serving as incentives as well as governors. The primary interest in research upon those conditions has been, however, with the governing effect. Motivation has been considered a secondary variable to be controlled and held constant rather than the major interest of the experimenter. Temperature, light, noise, rest periods, or the like may themselves influence the "will to work" or the effort expended by the worker. So far as possible, this aspect of their effect is eliminated from experimental study by special controls. Such control is not always complete, as we shall see, but it is important to keep this purpose in view as we survey the experimental information on these topics.

Noise

It is common to discuss the effects of noise under the heading of "distraction." We have deliberately avoided doing so because the experimental literature is limited to meaningless sounds or noise, and leaves out of account many other potential sources of distraction. In addition, it is safer to speak of the effects of noise than of the effects of distraction because the term distraction assumes that the noise has an effect which must still be demonstrated.

The emphasis which is given to noise abatement campaigns indicates a common assumption that noise is a hindrance to effective work.

The interest in controlling noise has, however, far outrun the factual evidence available. The sources of evidence which we have must be divided into two groups because of the distinct methods and conditions which were used. The first group consists of experiments upon productivity and effort of subjects working in a laboratory with high motivation and relatively short periods of exertion. The second group consists of scattered statistical analyses of production in the industrial situation, where it has been possible to introduce some control of the noise factor but not of many other variable conditions.

Laboratory Studies.—If we are interested in the effects of noise as a governing factor in performance, we wish to find its influence upon subjects who are well-motivated. Effects of noise upon the normal working rate of workers in the industrial setting may indicate either variations in motivation, or in capacity for performance, or both. The laboratory studies are important, therefore, in indicating whether or not noise has a *necessary* effect upon the performance of well-motivated workers.

The brevity of the working period in most of the laboratory studies gives little opportunity for the worker to become accustomed to the noise. In spite of this, it was soon discovered that the effect of intense and irregular sounds was surprisingly small. In many cases there would be a short period of reduced performance, but the recovery would be very rapid. Occasionally, the subject might actually accomplish *more* under noisy conditions than he did in quiet. Thus the influence of noise upon rate of performance in the laboratory can be considered as only a small and temporary one.

It is still possible that long periods of work under especially noisy conditions might show more effect upon output rates, due to some kind of cumulative influence of the sound. So far there is little experimental evidence upon these longer periods. In shorter periods the subjects apparently become accustomed to the sounds.

A number of different lines of evidence suggest, however, that this adjustment to the noise is attained at the expense of additional effort on the part of the subject. For some reason, the subject tends to adopt a certain level of output as his goal, rather than a constant level of effort.

In spite of the evidence that effort is increased during noisy conditions, there is some indication that the major effect of the noise upon effort occurs during the initial minutes of the noisy period, and the effort diminishes at the same time that the output is returning to normal. If this is true, the adjustment to the noise does not in-

volve a continued poor efficiency; instead, the total adjustment to the noise may be a return to normal efficiency as well as a return to normal working rate. The evidence upon this latter point is inconclusive as yet, so we shall examine it in order to see what remains to be done in solving the problem.

Let us look briefly at the evidence upon which these generalizations are based. The two main lines of evidence are studies of muscular

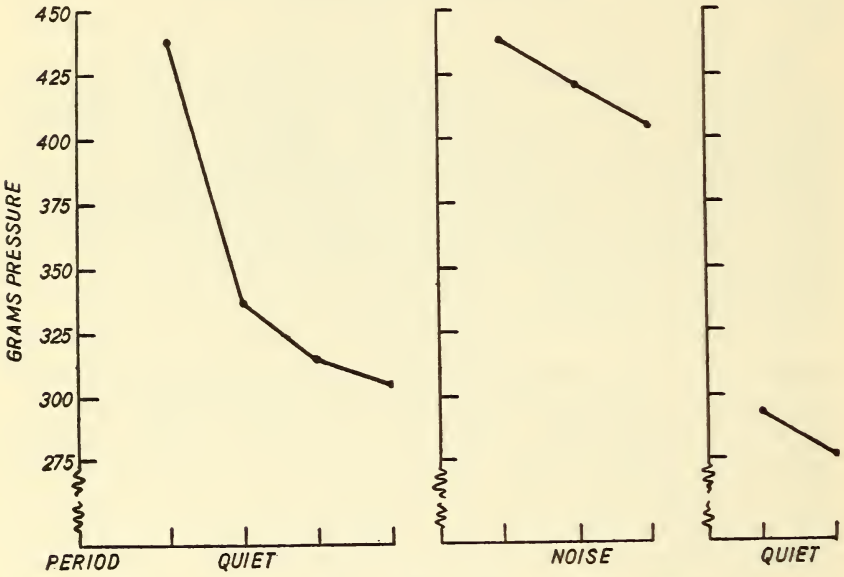


Figure 14. Key-Pressure in Work During Quiet and Noise

Points represent means of successive groups of 100 reactions each. (From data in Morgan, *op. cit.*, p. 44.)

tension during work, and metabolic cost. Morgan[1] recorded the pressure which the subjects exerted upon the reaction keys with which they performed their experimental task, and also recorded the breathing pattern during work. There were changes in both records during noise, the pressures increasing and the breathing becoming irregular in comparison to the working patterns of the quiet periods. The major increase in tension (pressure) occurs, however, at the beginning of the noise, and there is a definite fall in this tension in a relatively short time. (See Figure 14.) The change in breathing pattern during the period of noise is not clear, but it is suggested that the

[1] J. J. B. Morgan, The overcoming of distraction and other resistances, *Archives of Psychology*, 1916, No. 35, 5.

irregularities which appear during noise are due to the subject's "talking to himself" in order to avoid distraction. This may increase as time goes on. Whether this entails increased cost of the work or merely a change in the method of performing it cannot be decided.

Davis has measured muscle potentials directly in subjects who were at rest during quiet and noise, and for brief periods of work during noisy conditions.[2] Electrical potentials from the forearm increased when noise began, but soon dropped to their normal values again (Figure 15). In the second experiment which involved work

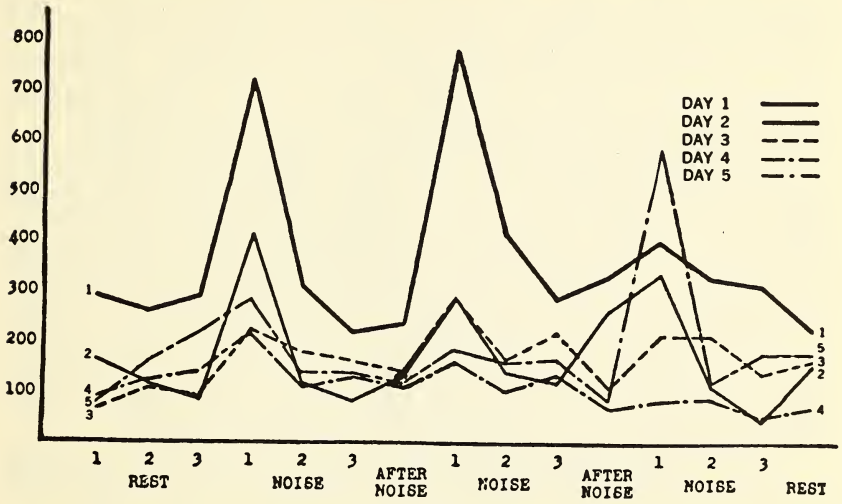


Figure 15. Average Magnitude of Action Potentials on Successive Days of Stimulation

(From Davis, *op. cit.*, 1935, p. 13.)

(difficult reading), the recorded period was only two minutes. During this time the muscular potentials increased. The period was too short, however, to determine the possibilities of adjustment to the noise.

The effect of noise upon the metabolic cost of work has been studied by Laird, for typing;[3] by Vernon and Warner;[4] and by Harmon,[5] for arithmetical calculation. Of these, the most carefully controlled

[2] R. C. Davis, The muscular tension reflex and two of its modifying conditions, Indiana University Publications, 1935, *Science Series* No. 3, and The relation of certain muscular action potentials to "mental work," *ibid.*, 1937, No. 5, 23-24.

[3] D. A. Laird, The measurement of the effects of noise upon working efficiency, *Journal of Industrial Hygiene*, 1927, *9*, 431-434.

[4] H. M. Vernon and C. G. Warner, Objective and subjective tests for noise, *Personnel Journal*, 1932-1933, *11*, 141-149.

[5] F. L. Harmon, The effects of noise upon certain psychological and physiological processes, *Archives of Psychology*, 1933, No. 147, *23*.

work was apparently that of Harmon. He found the metabolic cost of adding columns of figures was greater during noise than under quiet conditions. It should be remembered, however, that the increases in metabolism due to the work were not great in either case as compared to changes which are brought about by muscular work. Since the changes were small, and since there were only two subjects, it is doubtful how conclusive these results are, in spite of the many experimental precautions used by Harmon.

Several of these experiments have also included records of a measure of cost taken upon successive periods of exposure to the noise. Davis' experiments show a clear-cut decrease in the effect of noise upon muscle tensions, from day to day of his experiment. Harmon's subjects showed smaller and smaller effects of the noise upon oxygen consumption, as the observations were repeated several times.

Thus all the facts that are available suggest that there is little effect of noise either upon output or upon effort expenditure during work after the subject has become accustomed to the noise. At least this is one obvious interpretation of the known facts. Poffenberger has suggested another possible conclusion, namely, that the pattern of tensions *shifts* during the adjustment to the conditions of noise.[6] He leaves open the possibility that efficiency is still low, but that the change in pattern of effort is such that the effort is no longer so easily detected by the measuring devices used in these experiments.

Poffenberger's suggestion would be quite plausible if we were concerned only with Davis' results which involved muscular activity in the forearm alone. It is less satisfactory when applied to the other results we have described. Morgan's measure of tension involved muscles which were actually used in the activity under study, so that it seems less likely that tensions in other parts of the body would be substituted for those he measured as the process of adjustment to the noise went on. Similarly, if the oxygen consumption has any significance at all in these researches, it indicates the over-all muscular activity. As Harmon found, this over-all activity diminishes with time. The tension *pattern* might still be influenced by the noise, as Poffenberger suggests, but if it is just as economical as the pattern in quiet conditions, the pattern would be a minor consideration.

Such speculation avails little, of course, except to clarify the nature of the problem which still remains to be solved. We need a study of a greater range of factors in the cost of work. A more thorough

[6] A. T. Poffenberger, *Principles of Applied Psychology*, New York, D. Appleton-Century Co., Inc., 1942, p. 138.

study of bodily tension, including potentials from various portions of the body, and the skin resistance index of general tension, would be important contributions to this topic. We also need a more complete survey of the variations in these effects with varying durations of the noise, both longer experimental sessions and longer series of sessions, in order to study more completely the process of adjustment to the sounds.

All that we can conclude from the facts now before us is evidently that meaningless, intense noises are likely to reduce output slightly when they first begin, and, for that matter, when they first cease. During that same time the worker puts forth greater effort and compensates to a certain extent for the influence of the stimuli.

Field Studies.—If the rapid adjustment which is found in laboratory studies occurs also in the normal working situation, and if it continues in the same direction, we should expect that noise control would have little effect in industry. There is, however, the real possibility that such an assumption is unjustified, that the adjustment follows a different course, and that other accumulative effects of noise appear. Direct study of factory productivity, with reasonable control of conditions, is therefore a requirement. Although controls are difficult to obtain under practical conditions, nevertheless such studies can be made with present techniques. The study of *efficiency* under noisy conditions in the field is, of course, a more difficult problem from a methodological as well as practical point of view.

In spite of the need for field studies of production, and in spite of the apparent feasibility of such investigations, there are few significant findings available. In a recent review of this whole topic, Berrien notes that most of the articles dealing with these studies are written for the layman and give little indication of the procedures used. Many cannot be evaluated for this reason. He concludes further "Only one field study in this country has been completed with sufficient control and covering a sufficient time to place confidence in the results." [7] Even in this study the statistical significance of the findings has not been reported.

The investigation to which Berrien refers in the above quotation (carried out by the Aetna Life Insurance Company) and a study of weavers carried out by Weston and Adams in Great Britain [8] repre-

[7] F. K. Berrien, The effects of noise, *Psychological Bulletin,* 1946, *43,* 141-161.
[8] H. C. Weston and S. Adams, The effects of noise on the performance of weavers, *Industrial Health Research Board,* 1932, Report No. 65, pp. 38-62.

sent the small store of information which is even partially interpretable. The studies had similar results, and similar criticisms apply to the statistical interpretation.

For details of the field studies we refer to Berrien's excellent review. We need only say here that the results are consistent in indicating increased output when noise is reduced, even when the workers are accustomed to the noise. It appears that the adjustment to noise is never complete. The laboratory investigations therefore need to be extended in order to discover the source of the discrepancies. Perhaps the short duration of the work period in the laboratory has been confined to an initial period in which the subject changes his response to the noise very rapidly. As a result, the ease of adjustment has been exaggerated. There is also the possibility that noise operates primarily as a motivational factor in the normal situation, a phenomenon which does not appear in the laboratory because of the completely different goal of the subject. The question is whether a reduction in noise, serving only as an added incentive to effort, could continue to be effective over a period of months or years, as indicated by the field studies.

Very little is known of the effects of more meaningful sounds, such as conversation or radio programs. This lack is due in part to the great difficulty an experimenter would find in providing uniform experimental conditions. A repetition of the same conversations or programs would change their effect, while a change in content of the "distracting" material would provide extremely variable conditions for the experiment. It is also doubtful whether an artificially arranged or "staged" situation would have any similarity in effect to a similar distracting situation as it naturally occurs. At any rate, an experiment upon such meaningful sounds would require much more data in order to control some of these additional variables by means of statistical analysis. Previous researchers probably have also felt that it would be better to clear up the relatively simpler problem of noise before attempting these more difficult studies.

Music.—The frequently mentioned benefits of music as an accompaniment of work probably belong more clearly under the topic of incentive and interest in the work. The music may provide interest in an otherwise routine and dull task, or it may be taken by the worker as an evidence that management is interested in providing pleasant conditions of work. Wyatt and Langdon investigated the effect of music as a part of their larger studies of boredom which will be mentioned again in other parts of this book. They concluded that

the "effect seemed to be directly related to the amount of boredom experienced by the different individuals." [9]

That the effect of the music depends upon the "set" of the worker is also indicated by an experiment of Baker's.[10] He was able to show that the music could increase or decrease performance, depending upon whether the subject expected the music to be an aid or a hindrance to his work.

Temperature and Ventilation

Although the layman is often unaware of it, the problems of temperature regulation and the problems of ventilation for effective working conditions are usually identical. It is only under exceptionl conditions that there is any danger of oxygen deficiency or too much carbon dioxide due to inadequate ventilation. Many years ago it was conclusively demonstrated that the effects of bad air were attributable to faulty temperature regulation. In fact, "fresh air" is no more beneficial than "stale air" if the fresh air is not properly controlled with respect to temperature and the other factors which regulate the heat loss from the body. Naturally, we are excepting from this statement special industrial conditions in which noxious gases, dust, or other irritants are a danger. In the absence of these factors, even in a tightly closed room containing several people, it would take a long time to reduce the oxygen or increase the carbon dioxide to the danger point. Long before that, discomforts due to overheating are likely to set in. Thus the two topics which are included in this section comprise only a single common problem for efficiency in work.

Because of the interest of the medical profession and physiologists in the control of bodily temperature, and the interest of engineers in problems related to the design of heating and cooling equipment, a great deal of research has been devoted to the bodily changes which result from changes in temperature conditions of the environment. The American Society of Heating and Ventilating Engineers has supported a long series of these experiments. The Pierce Laboratory of Hygiene at Yale University has been another productive source of information on the physiological and biophysical sides of the problem.

This is not the place to attempt even a brief summary of the detailed physiological and physical facts and problems involved. We

[9] S. Wyatt and J. N. Langdon (assisted by F. G. L. Stock), Fatigue and boredom in repetitive work, Industrial Health Research Board (Great Britain), 1938, Report No. 77, p. 73.

[10] K. Baker, Pre-experimental set in distraction experiments, *Journal of General Psychology*, 1937, *16*, 471-486.

shall merely touch briefly upon a few facts which should be kept in mind when we turn to the psychological problems involved.

Let us consider first the physical factors involved. Most of us are aware of the importance of humidity and air movement, as well as of air temperature, as determinants of comfort and bodily temperature regulation. We often fail to realize, however, that another physical factor is also important. This is the factor of *radiant* heat loss or gain from our surroundings. The human skin is an efficient radiator or absorber of heat energy; in fact, it is nearly what the physicist calls a "black body" so far as heat energy is concerned. As a result, we can gain or lose heat from surrounding solid objects independently of the temperature of the air itself.

Because of radiation it is quite possible to remain comfortable in indoor clothing with the air temperature only 40° F. or 50° F. so long as the heat loss to the air is balanced by a lower rate of absorption by other objects. As an example, one engineer constructed a small room of polished copper. Because the walls absorbed so little of the heat given off by the body through radiation, the air temperature had to be quite low if the occupant was to feel comfortable. This factor is responsible for the fact that some rooms feel so much cooler at (say) 70° F. than do others. A large cold wall or a large window surface not only cools the air so that the room is harder to maintain at 70° F., but also absorbs radiant heat.

In order to specify the conditions which are satisfactory for a given type of work, therefore, it is necessary to specify four factors; air temperature, humidity, rate of air movement, and radiation (usually expressed as the average temperature of the surrounding solid objects). Statements that "ideal" conditions consist of 68° F. and 50 per cent relative humidity (as we are often told) are inadequate because these other conditions are not described, to say nothing of the nature of the evidence upon which such recommendations are based.

On what basis should standards of temperature and ventilation conditions be determined? Let us consider as briefly as possible the possibilities which have so far been investigated.

1. Effects upon physiological processes such as heart rate, blood distribution, perspiration, and the like. These measures are better suited to finding the extreme range of temperature conditions which can be borne by the body rather than the optimal conditions. For example, it is possible to specify the conditions which will produce a dangerous burden upon the heart, which will disturb the water

balance, or the like. There is, however, a wide range of temperature conditions within which the bodily changes are so small that we can attribute little importance to them.

2. Effects upon health and disease. A number of investigators have studied the effects of weather and climate upon the rate of incidence of disease and upon the death rate.[11] It has been more difficult to obtain significant data upon the influence of indoor conditions. Here it is necessary to deal with relatively small experimental groups in a situation in which it is difficult to eliminate the influence of many other important variables, such as difference in exposure, differences in conditions outside the experimental room, and the like. While there are many theories about the influence of temperature conditions upon health, there is little well-founded information here.

3. Influence of temperature conditions upon health, fertility, and activity of animals bred and raised under controlled conditions. While these studies serve to indicate that temperature *is* a factor worth considering, and may lead to the development of techniques which can later be applied to man, it is obvious that such observations could not establish optimal levels for man. As examples of this type of work we may refer to Mills' studies, in which he found that rats reared in a cool room (66° F. to 70° F.) grew faster and were more resistant to disease than those reared in temperatures of 89° F. to 92° F.; 60-70 per cent RH.[12]

4. Industrial performance and tested activities as influenced by temperature conditions. The Industrial Health Research Board of Great Britain has published a number of reports which show the influence of seasonal temperature changes and different methods of ventilation upon output in industrial activity. For heavy work, the main problem is naturally to keep the temperatures low enough to permit easy loss of the excess heat developed in the body.

For lighter activities, it has been difficult to discover the effects of temperature, if, indeed, they exist. The New York State Commission on Ventilation performed the most elaborate experiments upon this problem, using a variety of test activities, such as speed of naming colors, cancellation, addition and multiplication, and typewriting, along with tasks requiring heavy muscular exertion.[13] They were

[11] C. A. Mills, *Medical Climatology,* 1939. Ellsworth Huntington, Weather and Health, *Bulletin of the National Research Council,* 1930, No. 75, 1-161.

[12] C. A. Mills and C. Ogle, Control of body heat loss through radiant means, *Transactions of the American Society of Heating and Ventilating Engineers,* 1938, *44,* 167-178.

[13] *Ventilation,* Report of the N. Y. State Commission on Ventilation, 1923.

able to demonstrate that the influence upon muscular exertion was not a function of the "freshness" of the air, but of temperature conditions alone. The "mental" tasks, however, showed no effect of either temperature or freshness of the air. In fact, the subjects performed as well when the air was 86° F. and the humidity was 80 per cent as they did at 68° F. and 50 per cent R.H. When the subjects were left free to choose their own activity under the various conditions, there was no clear tendency to choose the easier tasks more frequently under warm conditions than there was under the more normal conditions.

One possible interpretation of these results is that the subjects were "put on their mettle," by the experiment. Perhaps, then, they compensated for the effects of adverse temperature conditions by increased effort, just as they did under the influence of noise. Against this suggestion it might be argued that under the "normal" conditions they were supposedly working nearly as hard as their capacities to perform, so that any reduction in capacity when the temperature increased should have made itself evident in the test scores. The same argument, however, could just as well be applied to the experiments on noise, where, as we have seen, it is not borne out by the facts. It is likely that it would not apply to the effects of temperature either. Apparently it is not possible to obtain maximal performance on mental tests of this kind. There is always some additional effort possible which may be brought out by adverse conditions of work—noise, or temperature, or others still to be discussed.

Another possible interpretation is that temperature does not, in fact, influence mental activity very much. This is quite possible, since the bodily heat production involved in such tasks is so small. The only possible way to decide the question is by further experimentation which takes *effort* as well as performance into account, using techniques similar to those employed for noise. This has not been carried out so far, and it is, therefore, impossible to determine standards for temperature conditions upon the basis of performance until this has been done.

5. Reports of comfort. It appears from the above summary of attempts to find some sensitive method of determining optimal temperature conditions that the only method now available for the purpose must be based upon the comfort of the subject, as perceived and reported by him. In many ways this also is subject to criticism, and we should prefer something better. We may choose temperatures which are too warm for our ultimate good just as we may eat more

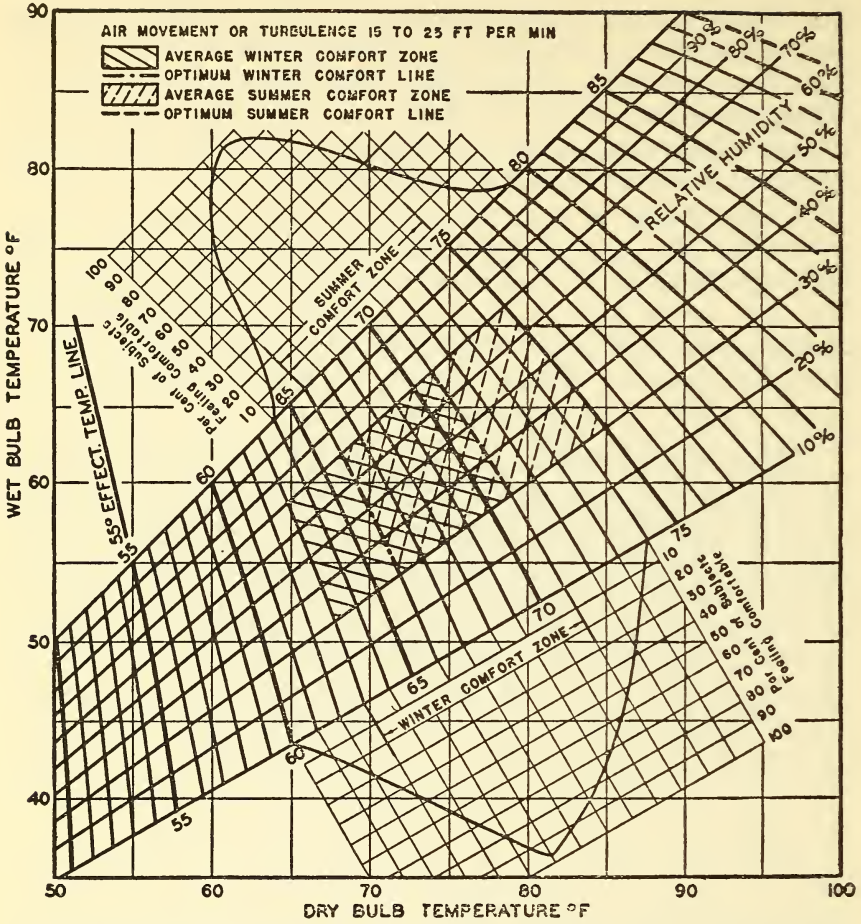
than is good for us, or too much or too little of certain foods. Nothing can be done about this, however, until the questions discussed above have been more adequately studied. It is possible, also, that rated feelings of warmth or cold are subject to unreliability, as are other ratings. They are probably affected by suggestion; they will also be influenced by adaptation and acclimatization.

In many studies in which the subjects report their degree of comfort, there is no specification of the conditions from the point of view of radiation. The other three important factors are usually reported. In the examples shown in Table 10 most of the results are given in terms of "Effective Temperature." It is necessary to understand this term before considering the results. The following is a rough statement of the meaning of the term.

Effective Temperature.—A scale in wide use by heating and air-conditioning engineers, which shows the combined effect of air temperature and humidity for a specified rate of air movement. It was established by subjects who compared the warmth of two rooms of different humidity and temperature. Temperatures were adjusted until the rooms felt equally warm while the relative humidities remained fixed and different. Thus 66° F. effective temperature means that the conditions would feel the same as a room at 66° F. dry bulb temperature and 100 per cent relative humidity. Knowing the wet-bulb and dry-bulb temperatures in a room, it is possible to determine the effective temperature equivalent from a chart (Figure 16). In the examples given, air movement is always at a minimum.

The results indicate that it is possible to suit a large proportion of the occupants of any building if their work is all of the same general sort, if seasonal variations in requirements are taken into account, and if variations in climate in various regions are also considered. There are individual differences, of course, but the high percentages of subjects who are comfortable at the optimal level indicates that the amount of severe discomfort would be quite small.

In establishing the optimal level in a given climate and for a given kind of activity, the votes should, of course, be obtained in a systematic and controlled manner. If possible, it would be well to keep the subjects in ignorance as to the thermometer readings at the time the opinions are being taken, in order to prevent biases from entering the judgments. It would also be well to record the opinions after some period of time under a given condition. Those who have become accustomed to a given level may take some time to become accustomed to new conditions.



Note.—Both summer and winter comfort zones apply to inhabitants of the United States only. Application of winter comfort line is further limited to rooms heated by central station systems of the convection type. The line does not apply to rooms heated by radiant methods. Application of summer comfort line is limited to homes, offices and the like, where the occupants become fully adapted to the artificial air conditions. The line does not apply to theaters, department stores, and the like where the exposure is less than 3 hours. The optimum summer comfort line shown pertains to Pittsburgh and to other cities in the northern portion of the United States and Southern Canada, and at elevations not in excess of 1000 ft above sea level. An increase of one deg ET should be made approximately per 5 deg reduction in north latitude.

Figure 16. American Society of Heating and Ventilating Engineers Comfort Chart for Still Air

Effective temperature may be determined from dry-bulb temperature and relative humidity (or wet-bulb) reading. Effective temperature lines run from upper left to lower right.

(From *Heating Ventilating Air Conditioning Guide* 1947. Reprinted by permission.)

TABLE 10

Zones of Maximum Comfort

Reference	Conditions	Temperature Scale	Maximum Comfort Reports	Per Cent of Comfortable Reports	Range over 50 Per Cent
Yaglou and Drinker (1929) [14]	Clothed; sedentary; adapted ½ to 1½ hrs.; air movement 15-25 ft./sec.	Effective temperature	(Winter) 66 (Summer) 71	97 * 98	63-71 66-75
Fleischer (1939) [15]	Clothed; light work; standing	Effective temperature	77.5		
Houghten (1937) [16]	Clothed; summer. (*Test Rooms*)	Effective temperature			
	Pittsburgh		72.5	94 **	68-74.5
	Toronto		70	80	?-71.5
	Boston		70.5	69	69-72
	Texas		73	87	70-74.5
Houghten (1939) [17]	Office workers; summer.				
	Washington	Effective temperature	71	95 **	?-75
	San Antonio		71.5	92	68.3-75
	Minneapolis		70	75	67-73.5
Bedford (1937) [18]	Light factory work; winter.	Effective temperature			Range of Comfort
	Great Britain	Equivalent temperature	60.8 ***		53.7-67.8 ***
		Dry bulb	62.3		52.8-71.7
			64.7		55.8-73.7

* Per cent comfortably cool, comfortable, or comfortably warm. ** Per cent comfortable.

*** Calculated from regression equation of comfort scale against temperature. Maximum taken at point where average predicted comfort vote would be "comfortable," and range taken from average comfort vote of "comfortably cool" to "comfortably warm."

[14] C. P. Yaglou and P. Drinker, The summer comfort zone: climate and clothing, *Transactions of the American Society of Heating and Ventilating Engineers*, 1929, *35*, 269-286. See also the *Journal of Industrial Hygiene*, 1928, *10*, 350-363.

[15] W. L. Fleischer, A. E. Stacey, Jr., F. C. Houghten, and M. B. Ferderber, Air conditioning in industry, *Heating, Piping, and Air Conditioning*, 1939, *11*, 107-111, 191-194, 255-268. See also *Transactions of the American Society of Heating and Ventilating Engineers*, 1939, *45*, 59-110.

[16] F. C. Houghten, F. E. Giesecker, C. Tasker, and C. Gutberlet, Cooling requirements for summer comfort air conditioning, *Transactions of the American Society of Heating and Ventilating Engineers*, 1937, *43*, 145-184.

[17] F. C. Houghten, C. Gutberlet, and A. A. Rosenberg, Summer cooling requirements in Washington, D. C., and other metropolitan districts, *Heating, Piping, and Air Conditioning*, 1939, *11*, 587-591. See also *Transactions of the American Society of Heating and Ventilating Engineers*, 1939, *45*, 577-590.

[18] T. Bedford, The warmth factor in comfort at work, a physiological study of heating and ventilation, Industrial Health Research Board (Great Britain), Report No. 76, 1936, iv, 102.

Such votes are useful as a stopgap for use until other techniques can give us more satisfactory methods of determining optimal temperature levels.

Lighting and Other Factors in Ocular Work

Campaigns for "better light, better sight," and the new developments in lighting methods, have in recent years increased the attention which is given to the conditions of lighting for work such as reading, drafting, visual inspection, weaving, sewing, and many other similar activities. There are a number of interrelated questions involved in finding the optimal conditions for these tasks. There are problems of intensity, distribution and color of the illumination, and other factors, such as color of materials, which play a part. Intensity cannot be changed without producing some effects upon glare and distribution of light, unless the lighting fixtures are also modified. Other interrelations hold for other combinations of these factors. While we shall consider these aspects separately in the following paragraphs, it is only an artificial separation, and the reader must keep these relationships in mind.

Intensity of light.—Assuming optimal distribution and constant color of light, how bright should the light be for a given task? This question again raises the problem of the cost of work to the visual mechanism—the ocular fatigue involved. There are still differences of opinion concerning the correct levels of illumination to be adopted, due to the differences in evaluation of the measures of ocular fatigue.

Two distinct points of view are represented by Tinker and Luckiesh. Tinker points out that increases in illumination bring about increases in performance up to a certain point, as shown in Table 11. Beyond a certain level, however, no further advantage is gained by additional brightness of light.[19] Tinker has collected the results of a variety of studies to indicate the critical level for different tasks. These levels are shown in Table 12. Tinker does not believe there is sufficient evidence for any advantages to be gained by brightnesses very much above these values.[20]

Luckiesh, on the other hand, takes the position that eyestrain may be decreased by brighter lighting, even if there is no increase in per-

---

[19] M. A. Tinker, Illumination standards for effective and comfortable vision, *Journal of Consulting Psychology,* 1939, *3,* 11-20.

[20] Tinker makes recommendations for practical lighting somewhat above the critical levels, *e.g.,* about fifteen foot-candles for ordinary schoolroom tasks, allowing for a safety factor. See M. A. Tinker, Illumination standards, *American Journal of Public Health,* 1946, *36,* No. 9, 963-973.

formance to indicate the benefit. It is conceivable, of course, that there are other limiting factors in ocular work which would establish a ceiling on performance and prevent indefinite improvement of out-

TABLE 11

EFFECTS OF INCREASED ILLUMINATION UPON INDUSTRIAL OUTPUT *

Type of Factory	Initial Level (*Foot-Candles*)	New Level (*Foot-Candles*)	Per Cent of Increase in Production
Electrical	3.8	11.4	8.5
Piston Ring	1.2	6.5	13.0
" "	"	9.0	17.9
" "	"	14.0	25.8
Roller Bearing	5.0	6.0	4.0
" "	"	13.0	8.0
" "	"	20.0	12.5

* After Troland (table assembled by Tinker, *op. cit.*, 1939, p. 17): To present a concrete idea of the meaning of the foot-candle unit of illumination, a 100-watt bulb in a bridge-lamp shade will give 10 foot-candles at a distance of 52 inches, and 25 foot-candles at 34 inches. There is, however, variation with the size and color of shade, brightness of walls, and the like.

TABLE 12

CRITICAL LEVELS OF ILLUMINATION †
BEYOND WHICH THERE IS LITTLE CHANGE IN OUTPUT

Task	Critical Level (*Foot-Candles*)
Reading 10-point type (speed)	3
Reading 10-point type (fatigue, blurring)	3
Reading 12-point type (speed)	Between 1 and 10
Reading performance (school)	4-6
Educ. achievement in school	4-6
Computing arithmetic problems	Less than 9.6
Sorting mail	8-10
Setting 6-point type by hand	13-25
Time to thread a needle	30

† Tinker, *op. cit.*, 1939, p. 13.

put. Thus, in reading, we cannot read faster than we are able to comprehend. Luckiesh's argument would be that output alone cannot be used to establish the optimal levels of illumination, and this seems a reasonable argument. The final decision depends, however, upon our ability to measure fatigue in ocular work by methods independent of the output.

In a series of studies in the General Electric Company laboratories at Nela Park, Luckiesh and his collaborators have investigated the effect of illumination upon a number of proposed measures of ocular fatigue. The choice of these measures is based upon the following reasoning: (1) these tests involve the ocular mechanisms (accommodation, etc.) or other mechanisms which are involved in the total activity of reading (*e.g.*, muscular tension) and (2) these measures change under conditions which are obviously bad (the subject wears incorrect lenses or works under a very bad glare).

A summary of the effect of intensity upon these measures is given in Table 13.

TABLE 13 *

INFLUENCE OF LEVEL OF ILLUMINATION

Type of Test	Intensity of Illumination (Foot-Candles) 1	10	100
Visual acuity	100 per cent	130 per cent	170 per cent
Contrast sensitivity	100 per cent	280 per cent	450 per cent
Muscular tension (Key pressure) while reading	63 grams	54 grams	43 grams
Frequency of blinking (Change after reading one hour)	100 per cent	77 per cent	65 per cent
Decrease in heart rate while reading one hour	10 per cent		2 per cent
Decrease in convergence reserve of ocular muscles after reading one hour	23 per cent		7 per cent

* M. Luckiesh, *Light, Vision, and Seeing,* New York, D. Van Nostrand Co., Inc., 1944, p. 200.

It will be noted that two of the measures listed in Table 13 are measures of capacity for discrimination (acuity and contrast sensitivity). They do not indicate the cost of performing normal tasks where the discriminations required are far above the limits of vision, unless we make the assumption that all work is made easier if the limits of performance are raised. Such an assumption has been questioned, however, so that the effect of illumination upon capacity alone cannot be conclusive in settling the problem. Other measures of capacity have also shown similar effects of raising the intensity of illumination. Ferree and Rand, for example, have found increased speed of discrimination and speed of focusing, the speeds still increasing up to 100 foot-candles.[21] In all cases, acuity or speed increase

[21] C. E. Ferree and G. Rand, Intensity of light and speed of vision studied with special reference to industrial situations, *Transactions of the Illuminating Engineering Society,* 1927, *22,* 79-110.

rapidly as the light passes from zero through the lower ranges of intensities, and more and more slowly as higher levels of intensity are reached. In Luckiesh's data the increases are proportional to the *ratio* of increase of light rather than to the absolute increase. Thus the gain from additional light energy is relatively small, once we pass beyond 10-20 foot-candles. Since this phenomenon of diminishing returns applies to the other measures of cost, it is used by Tinker as an additional argument against the desirability of high intensities.

The other measures (tension, heart rate, convergence reserve, blinking) are considered as indices of the strain which ocular work puts upon the organism as a whole or upon the ocular musculature. These results have been questioned by Tinker and others, not only because of the diminishing returns already mentioned, but also because the measures are regarded as of doubtful validity or reliability. Space will not permit an analysis of this question here, but it appears that at least the rate of blinking and of the heartbeat have little value as indicators of fatigue in ocular work.[22] In addition, it should be mentioned that Tinker has experimented with a further test of fatigue which Ferree and Rand applied with success to problems of light distribution and color. Tinker could find no change in this index beyond about 3 foot-candles.[23]

Ferree and Rand approached the question of intensity of light from the point of view of individual differences. They had found that additional light has more marked advantages for the older person who has lost part of his power of accommodation.[24] Acuity showed greater increases for these subjects than it did for younger, normal subjects. Their later studies of individual differences related to preferences in light rather than to functional tests.[25] Since these preferences are likely to be a function of habits and customs of the subjects, it is not likely that they will solve the problem before us.[26]

---

[22] R. A. McFarland, C. A. Knehr, and C. Berens, Metabolism and pulse rate as related to reading under high and low levels of illumination, *Journal of Experimental Psychology,* 1939, *25,* 65-75. R. A. McFarland, A. H. Holway, and L. M. Hurvich, *Studies of Visual Fatigue,* 1942 (Boston). M. E. Bitterman, Heart rate and frequency of blinking as indices of visual efficiency, *Journal of Experimental Psychology,* 1945, *35,* 279-292.

[23] M. A. Tinker, The effect of illumination intensities upon the speed of perception and upon fatigue in reading, *Journal of Educational Psychology,* 1939, *30,* 561-571.

[24] C. E. Ferree and G. Rand, The effect of increase of intensity of light on the visual acuity of presbyopic and non-presbyopic eyes, *Transactions of the Illuminating Engineering Society,* 1934, *19,* 296-313.

[25] C. E. Ferree and G. Rand, Good working conditions for eyes, *Personnel Journal,* 1937, *15,* 333-340.

[26] Tinker has stressed the importance of the adaptation of the subject in evaluating the results of a preference study (M. A. Tinker, Effect of visual adaptation

Of course the same might be said of preferences for temperature conditions, but here we must recognize the existence of ocular strain as the central problem to be solved. In the field of temperature there is no clearly recognized strain due to temperature unless the conditions are extreme.

To conclude, there is little indication in this set of contradictory findings that any great benefits are to be expected from levels of light above 10 foot-candles for ordinary reading, with somewhat higher levels for more exacting tasks such as fine sewing or inspection work. With modern lamps and lighting fixtures, there is probably little danger of inadequate intensity of light, although that was not the condition a few years ago. In special tasks, however, greater attention to this problem may still be necessary. There is also no evidence that high intensities are harmful, provided that the factor of glare is eliminated.

Glare and Distribution of Light.—Glare is probably a more important factor in ocular efficiency than intensity of light, at least under average conditions. We can discuss glare more briefly, however, since there is no disagreement about the problem.[27] Bright spots in the field of vision make the seeing of details more difficult, and all the various indices of ocular fatigue show the effect of these conditions. This is the general principle, which has a number of implications.

---

upon intensity of illumination preferred for reading with direct lighting, *Journal of Applied Psychology,* 1945, *29,* 471-476). His experiment, dealing with localized lighting, did not, however, indicate any marked effect of the adaptation level upon choice. The results were about the same whether the subject was adapted to twenty or to fifty foot-candles before making his choice. When choosing more generally distributed illumination, however, Tinker found a more marked effect of the adaptation level.

The adaptation to which Tinker refers includes only one factor which may play a part in determining preferences for intensity levels. In addition to this factor in the immediate history of the ocular mechanism, there is probably a more general kind of "growing accustomed" to certain general levels of lighting for reading or other ocular tasks. This factor could be controlled only by finding out the habits of the subjects, the type of lamp which they ordinarily use, the customary position when reading, and the like. If sufficient data of this kind were collected, then the factor of habituation might be controlled by statistical methods.

[27] The measures of fatigue and capacity which have been used in the study of glare are the same as those employed in studying the effect of intensity. Reliability and validity are no greater in one case than in the other, but since there is greater agreement among various measures and different investigators, the importance of the factor of glare is usually accepted without question. For example, Ferree and Rand's *li* test, which showed no advantage of high levels of intensity, is reported to show very consistent effects of bright light sources in the field of vision.

It should be noted, however, that until these measures are more fully established and validated, we cannot say *how* serious it is to have a given bright source in the field of vision. We can be sure it has some harmful effect, but how great this effect would be in comparison with other effects, such as the fatigue due to reading very fine print, is very difficult to say.

Lighting fixtures should not be visible when the subject is working. Shading the bulbs is not enough, because the shade rim itself will be a bright spot contrasting with surrounding darker objects. Therefore the fixtures should be removed from the field of vision altogether. "Indirect" systems of lighting are not always satisfactory for this purpose, because the ceiling or walls which reflect the light may themselves become bright spots if sufficient intensity is supplied at the working place. It should also be noted that bright spots may be produced by highly polished objects as well as by lighting fixtures.

While there should be no bright areas in the field around the work, there should also be no very dark areas. If a part of a room is very dark, lighter objects will contrast with the dark areas and produce the effects of still brighter objects. For these reasons it is usually recommended that the room be supplied with a moderate level of even illumination, with extra light falling upon the work from fixtures which are above and behind the worker.

Color of Illumination.—Among monochromatic (single wavelength) lights, the eye is more sensitive to the middle or yellow part of the spectrum. Even when the illumination is constant in apparent brightness, and varies only in hue, Ferree and Rand found that acuity was greater under yellow or amber light than it was under any other colored light. They also found a difference in fatigue (*li* test) produced by reading under the various colored lights.[28] Among mixed lights, it is found that daylight is more effective than tungsten lamps. It must be remembered, however, that filters for tungsten lamps which duplicate the spectral distribution of sunlight also reduce the intensity of the light. This means, of course, a greater expense in maintaining the level of intensity required. Ferree and Rand report that some so-called "daylight" filters do not produce any advantage over ordinary tungsten lamps.

Working Materials and Conditions.—Ease of seeing is a function of the amount of contrast between details to be discriminated and their background. This contrast is enhanced by adequate illumination, but it can also be controlled by the manufacturer of books, the designer of equipment, the painter of machinery. Paper should be as white as possible, and have a surface which does not produce mirror-like reflection. Newspapers and catalogues are often fatiguing

[28] C. E. Ferree and G. Rand, Visibility of objects as affected by color and composition of light: Part I, *Personnel Journal,* 1931, *9,* 475-492; Part II, *ibid.,* 1931, *10,* 108-124.

because of the quality of the paper. Dark parts or tools can be seen more readily if they are provided with a light background. It has been found helpful in some industrial operations to work on a ground glass plate through which diffuse light shines from below.

These are a few examples of the many devices which may increase ease of discrimination without raising the level of illumination, and often at very small expense. Research is also providing book manufacturers with much information upon type faces, margins, leading, etc., all of which affect the ease of reading. On some points there is controversy, due to the same difficulties which were met in determining optimal levels of illumination. Nevertheless there are enough facts established to provide a sound beginning. Until measures of fatigue are more adequately validated, it would probably be best to follow the suggestions of Paterson and Tinker which are based upon performance tests.[29]

There are many other problems of visibility which can be settled only by special psychological experiments. The optimum arrangement and design of automobile headlights to provide maximum distance of perception at night, the placement and design of highway signs, and other similar information, are of most immediate interest to the engineer or highway official, since the general public cannot put such information into practice. The applied psychologist needs to be aware of the existence of studies of this kind and be prepared to recommend that specialists in this kind of work be called upon to study new problems as they arise.

As one example of this kind of work, we may refer to the recently published work of Berger upon the visibility of numbers.[30] He has studied the shape and proportion of numerals, and has designed a set which are more visible than those now used on license plates. In addition, all the numbers are equally visible. In this analysis, the effects of width of line, spacing of numbers, black on white versus white on black, the shape, and area of the numbers were all taken into account. In this manner it was possible to produce numerals of the same size which were about 36 per cent more legible than previous numerals which were used for comparison. White numerals on a black background were more easily recognized than black numerals on a white background, and slender luminous numerals were more effective at night than the numerals seen under reflected light.

---

[29] D. G. Paterson and M. A. Tinker, *How to Make Type Readable,* New York, Harper & Bros., 1940.

[30] C. Berger, Stroke width, form, and horizontal spacing of numerals as determinants of the threshold of recognition, *Journal of Applied Psychology,* 1944, *28,* 208-231, 336-346.

Hours of Work, Rest Periods, and Sleep

Some of the effects of these factors have appeared in laboratory studies of fatigue and efficiency, and have already received brief mention in this book. Their more general effects in the world of work outside the laboratory must now be surveyed. The scheduling of working time, which involves all the factors listed in the title of this section, has come in for considerable discussion and some research. It is a topic which becomes especially important during periods of high production requirements. Even in slack seasons when the labor supply is plentiful, there are problems of this kind which require research. For example, is it more efficient to employ many workers for very short periods, or fewer workers for longer periods?

The effects of the length of the working day, of the weekly hours of work, of formal programs of rest periods in industrial production, have all been investigated primarily in terms of their effect upon total production, with an occasional reference to "morale," turnover, and accident frequency. The effects of these factors upon efficiency, in our sense of the term, have either been altogether overlooked or deduced by commonsense analysis from production data. Some of the recommendations for optimal conditions, particularly in regard to rest periods, have not been based upon direct study in industry, but have been drawn from laboratory studies of work on the ergograph, or some other task whose resemblance to normal working activities is extremely remote.

We shall not have space for a detailed examination of the research on these topics. We shall try only to indicate its general character, and give some of the main conclusions which can be drawn from the group of researches now available.

Hours of Work.—Problems of wartime production during the first World War gave impetus to the study of optimal hours of work from the point of view of obtaining maximum total output per worker. During that war the British established a Health of Munitions Workers Committee which investigated the problem. Later, the Industrial Health Research Board continued the analysis, and drew conclusions which have been supported by more recent investigators. Most of these results are summarized by Vernon.[31]

Several of these studies were begun at a time when the weekly shifts in munitions factories were extremely long (often over seventy

---

[31] H. M. Vernon, *Industrial Fatigue and Efficiency,* New York, E. P. Dutton & Co., 1921, pp. viii, 263.

hours per week). It was found that a reduction of the weekly hours increased the average hourly rate of production. Often this increase in rate of work was so great that it more than balanced the decrease in number of hours worked. The effect was not so marked in those jobs where the rate of work was largely controlled by automatic machines. Even here, however, the loss in total output was not so great as the reduction in number of hours worked. This phenomenon is due to the fact that total output is a function of the actual working time of the individual (nominal time minus rest periods, interruptions, etc.). A smaller proportion of working time is lost from these interruptions in the shorter shifts.

The optimal length of a working week cannot be stated in general terms. As indicated above, the optimum, from the point of view of total production, varies with the nature of the job. In some of the jobs reported by Vernon, the largest output occurred at fifty-five hours (the lowest studied), in others the optimum was at the highest number of hours, or over seventy. It must be remembered that most studies of this kind involve the special motivational conditions of wartime. There is no guarantee that the same optimum would be found in peacetime, even for the same kind of work. The only way in which the optimum for a given job can be determined is by long-term trial.

Another difficulty with attempts to state optimal hours of work is that the optimum, from the point of view of *total* production, is not necessarily the optimum from other points of view. In some ways we might consider the average hourly *rate* of work as a better indicator of the efficiency of the individual worker. The hourly rate also shows variation with varying length of shifts, going down when the shifts are either very long or very short. The highest rate of work often appears, however, in a shift which is shorter than that which brings maximum total production.

As we have suggested before, however, the rate of work alone is no more satisfactory as an indicator of efficiency than is the total output. We might work at maximum speed if we worked only one hour per day. Nevertheless, total output must be given some consideration. Our ideal is a maximum ratio between the total output per week and the total cost of the work for the week. This point is probably not the point where we attain the maximum total output, nor is it the point where we achieve a maximum hourly rate of work. A point somewhere between would be the most likely optimum. At present, however, we have no information which would permit us to determine that optimal point for the hours of work, even for a

single job. Until information of that kind can be secured, and in times of pressure for high production, the point of maximum total output would appear to be the best choice.

The Industrial Fatigue Board studies provide some indirect clues as to the effect of changing hours of work upon *efficiency* of performance. When the length of the working week is reduced, it may take as long as eight months for the rate of production to reach a new equilibrium. When the week is first shortened, production may go along for some time with only slight increases in hourly rate. Then it gradually moves up to a new level. In some instances the first rise in production rate may overshoot its ultimate level for a time, with a subsequent reduction to the new stabilized level.

This gradual process of adjustment of the worker to the new schedule might be interpreted to mean that the worker, intentionally or unintentionally, divides up the quota of effort which he is willing to expend on the job. He spreads that quota over the working week. When the working week is shortened, he has to learn how to re-divide his quota of effort, gradually increasing the allotment to each hour and day until he is again expending the same total which he used before. This is a rough statement of the interpretation which Vernon gives. If it is true, it would mean that an increase in total production means an actual increase in efficiency. Since the cost is adjusted at approximately the same level as before, the increase in output is pure gain.

To follow the logic of this interpretation still further, we may assume that the worker is also limited in the rate at which he can expend effort at any one time. As the hours of work are shortened below the optimal point, he has more and more difficulty in expending the quota of effort which he is willing to give to the work. Consequently the total output goes down.

When the hours of work are increased, however, the process of adjustment is no longer gradual. There is an immediate reduction in hourly rate of work, although here again there may be an overshooting, with some gradual readjustment later. It would appear that the mechanism of this adjustment is different from, and probably more deliberate than, that which accompanies reductions in the hours of work.

We could probably build still further upon Vernon's theory of a quota of effort (which Vernon calls "energy"). For example, the effect of changing motivation might be described in terms of a reduced quota of effort which the worker is willing to devote to the job. In fact we shall make use of this conception in our later chap-

ters upon the special topic of motivation. Before building too far, however, we must remind ourselves that the whole theory is based upon a supposition which needs to be checked experimentally. This represents one of the important tasks awaiting the development of more complete indices of effort and the cost of work.

Another, more immediate, implication of the gradual adjustment to reduced hours of work concerns the practical problem of finding the optimal hours of work for a given job in terms of total output. Since there is no single optimum, as we have indicated before, each plant or industry must determine its own recommendations. These recommendations must not, however, be based upon brief periods of observation. Vernon's studies indicate that at least eight to twelve months must elapse in testing each length of working period. We might also suggest that motivational conditions should remain at least approximately constant during that time if the results are to be meaningful.

The slow adjusment to decreased hours and the rapid adjustment to increased hours indicates, according to Vernon, a serious disadvantage of irregular overtime work. He believes that a worker who is called upon occasionally to work for a ten-hour shift in place of a regular eight-hour day, is likely to work at a rate adapted to the ten-hour period, not only on the days when he knows he is to work longer, but all the time. For this reason it would be better to spread the overtime evenly throughout the week, rather than concentrate it in one or two days.

The comparative effects of night and day shifts have been discussed at some length on a common-sense basis. Very little accurate information is available, however. In many plants the workers on the night shift are not equivalent in skill to those on the day shift, since seniority may give a worker a right to choose daytime work. In some instances the supervision is also less adequate at night. For these reasons, comparison of production records would have little meaning. The answer to the question of whether a night shift is less efficient than a day shift is largely academic, in any event. When workers are employed on a night shift, they are employed because the plant must be kept in operation day and night, and there is no choice involved. A more pertinent question, therefore, would be: What factors in efficiency differ in the day and night shifts, and how can they be controlled in such a way as to increase the effectiveness of both?

Whenever factories are working "around the clock," there is one choice which can be made. This is the choice between assigning

workers permanently to a given shift, or using a system of rotation. The major advantage of the rotating shift is that it is fairer to all concerned, since few prefer to work at night. On the other hand, it takes some time for an individual to readjust his habits for night work. He may not be able to complete the adjustment before it becomes time to change to new working hours.

Vernon has also reported statistics bearing upon this problem. The output of a group of workers on the night shift differed little from the output of the same workers when working in the daytime. This finding is not conclusive, according to Vernon, because deleterious effects of the night work might carry over into the succeeding period of day work, the periods of alternation being rather short. In other cases he compared the output of workers who were alternating shifts, with their later output when they were working continuously in either a day shift or a night shift. The output on the continuous day shift was only slightly better than that for the alternating shifts. The group which changed to the continuous *night* shift, however, declined about 4 per cent in output.

Thus there seems to be but little evidence of any marked effect of the arrangement of shifts upon the output of the workers. The most important factor in the choice would therefore be the preference of the workers as a group—the choice of the arrangement which seems fair to the greatest number.

Rest Periods.—The basic question in a discussion of rest periods in industry does not involve a choice between the presence or absence of rest periods. The studies of Vernon and others indicate clearly that some form of rest period is taken by the worker even where there is no formal plan allowing for them. The problem consists rather of determining whether or not a formally authorized rest period is more effective than an unauthorized rest, and if so, what program of rest is most beneficial.

As in the case of hours of work, the principal criteria of the value of rest periods have been hourly rate and total output per day. The results of introducing regular rest periods vary from slight increases in hourly rate which are insufficient to offset the time lost in the rest period, to increases which are sufficient to increase the *total* output by substantial amounts. There are enough data to warrant the conclusion that nothing is lost by the use of regular rest periods, and considerable gain in output and "morale" may result.

It is to be expected that some arrangements of rest periods would be more effective than others, but there have been few experiments in

which the schedule of rest periods has been varied systematically in an industrial setting. This is understandable when we consider that it may take a long time for the full effects of the rest periods to be reliably established.

The laboratory studies of rest periods had indicated similar general principles for both muscular and mental work. Graf, for example, performed experiments upon rest periods in continuous arithmetical calculation (three-hour periods). He found that a given amount of time spent in rest was more effective if the rests were brief and frequent.[32] In different series of experiments, a total of twelve minutes of the three-hour period was divided into anywhere from two to eleven rest periods. In general, the larger number of short rest periods led to the greatest gains. Graf also found that the results were better if the length of the rest periods was gradually increased from the beginning to the end of the work-spell.

Even though we do not have very much direct confirmation from industrial studies, there seems to be general agreement that these principles derived from laboratory studies are applicable to everyday tasks. One instance of a systematic variation in rest periods is Richter's study of women working in a metalworking plant, which shows a pattern similar to that obtained by Graf for a laboratory task.[33]

One of the most elaborate studies of the effect of rest periods will not be analyzed in detail in this chapter because the results have more bearing upon the problem of motivation than they have upon the effects of rest periods as *governing* factors in the efficiency of work. These studies, generally known as the "Western Electric" or "Hawthorne" experiments, began as studies of the effects of rest periods but resulted in few conclusions which have a bearing upon that problem.[34] The experiments failed to control the motivational factors which must be held reasonably constant in an adequate study of the present problem. The motivational factors loomed so large that the investigators became more interested in studying them.

As our later analysis of the Hawthorne research will indicate, there is nothing in these results which contradicts the opinion that rest periods are beneficial, although some readers have mistakenly drawn this conclusion. In fact, that company proceeded to install rest

---

[32] O. Graf, Über die Wirkung mehrfacher Arbeitspausen bei geistiger Arbeit, Kraepelin's *Psychologische Arbeiten,* 1925, *9,* 1-69, and Die Arbeitspause in Theorie und Praxis, *ibid.,* 1927, *9,* 563-681.

[33] W. Richter, Leistungssteigerungen in der Blankschraubenfabrikation durch Einführung von Zwangspausen, *Industrielle Psychotechnik,* 1931, *8,* 129-146.

[34] F. J. Roethlisberger and W. J. Dickson, *Management and the Worker,* Cambridge, Harvard University Press, 1939.

periods in other departments even though they could not draw conclusions on their effects from the experimental studies.

It is true that rest periods often have as much (or more) effect upon the motivation of the worker as they do upon the relief of fatigue. Motivation itself is a factor in efficiency, however, although the way in which it affects efficiency may be somewhat different from the effects of other factors. Even if the effect of rest periods were due wholly to increased motivation of the worker rather than to relief of fatigue, there is still a problem of efficiency to be considered. It is true that most of our knowledge of the effect of rest periods concerns output rather than efficiency. This indicates a gap in our knowledge which must be filled. Until it is, we may assume that the increases in productivity of the worker under rest periods are a reflection of increased efficiency.

There have been occasional studies of mechanical efficiency as a function of rest periods. In one of these, concerned with machine operators in a foundry, productivity and efficiency were both increased by rest periods.[35] In other studies, output decrement has been reduced by the introduction of rest periods. The work curves of Graf, for intellectual work in the laboratory,[36] and of Richter for factory work[37] both indicate this flattening of the output curve with optimal arrangement of rest periods.

This information upon mechanical efficiency and upon work decrement covers only part of the factors involved in total efficiency, and only a few specific jobs. We also have, however, the fact that output increases on hourly paid jobs when rest periods are introduced. This should give us additional confidence in the efficiency of formal rest periods.[38]

Sleep.—The topic of sleep involves questions which are usually the personal concern of an individual rather than problems which properly belong in the field of industrial psychology. Nevertheless, in certain occupations, sleep needs become a serious concern and a matter of policy to the organization as a whole as well as to the individual worker. Such occupations are aviation, transport in general, and

---

[35] A. S. Borschtschewski, S. D. Gorkin, and B. J. Kustanowitsch, Rationalisierung des Arbeitsregimes beim maschinellen Formengiessen, *Arbeitphysiologie*, 1932-1933, *6*, 311-338.

[36] O. Graf, *op. cit.*

[37] W. Richter, *op. cit.*

[38] Little has been said about the statistical reliability of the results cited in this section. In most cases this is not given, and the investigators do not appear to recognize this need. Nevertheless there is sufficient uniformity of results to justify some confidence in the general conclusions which we have drawn.

many military occupations. Even if it does have some importance in certain phases of industrial psychology, however, the topic of sleep is so complex that we shall attempt to give no more than a brief summary of its present status.

Here, as in many other phases of applied science, the solution of the immediate practical problems awaits the solution of more fundamental problems. The immediate problems are, of course, such questions as these: "How much sleep does an individual need?" "How long can one go without sleep without endangering health or safety in operating a car or machine?" "How can the depth or 'goodness' of sleep be improved?" "How can insomnia be prevented or cured?"

These questions cannot be answered, however, until we know more about the nature of sleep itself, what function it performs for the organism, and what factors affect it. In order to develop reliable experimental techniques which would arrive at knowledge upon these fundamental problems, it is necessary to develop some adequate criteria for determining when an individual is asleep, and how deep his sleep is at any given time. There are so many borderline states between sleep and waking, and an individual's own reports upon sleep are so untrustworthy, that the development of adequate criteria is an essential step in the study.

As an example showing the importance of criteria of sleep, we may mention the question of the distribution of sleep during the hours of the night. It is frequently claimed that we sleep more deeply during the early hours than we do toward morning. There are a great many bodily functions which undergo changes correlated with sleep. Some of these, however, show the greatest effect during the early hours, while others show the greatest effect later on. Since we have no way at present of saying which criteria are more intimately related to the depth or soundness of sleep, we can take our choice of criteria and prove that the soundest sleep occurs at any period we wish. We shall not go into the details of these bodily functions since they are reviewed with considerable completeness in Kleitman's book on the subject.[39] At any rate, there is no evidence to support the superstition that "we do most of our sleeping" in a few hours, and that the remainder of the time is wasted.

We cannot state definite sleep requirements either for individuals or as group averages because there is as yet no way of measuring the

---

[39] N. Kleitman, *Sleep and Wakefulness,* Chicago, University of Chicago Press, 1939.

effects of loss of sleep in any meaningful way. In terms of quantifiable measures, even prolonged loss of sleep (three or four days) produces surprisingly little effect. Kleitman found that physiological measures involving circulation, excretion, and the like, showed almost no change at all. Psychological tests also showed little effect, unless the test called for prolonged effort and continuous attention during a considerable period of time. For brief periods, performance was just as good as that in control periods.

Johnson has stressed the point of view that the nature of these results is a function of the tests which are used.[40] In order to get reliable and quantifiable tests, the tasks have to be rather simple and routine in character. Subjects who have been without sleep for prolonged periods may show many disturbances, such as hallucinations, irritability, delusions, and the like. Yet they perform well on the routine tasks in the tests. These findings are very similar to those of experiments upon the effects of alcohol, and so Johnson suggests that sleep deprivation has effects very similar to those of a narcotic. Disturbances very evidently occur, but their maximum effect involves functions which have not yet been adequately brought under the control of psychological tests.

This conclusion of Johnson's has a significance beyond the topic of sleep alone. Evidently tests which cannot detect the effects of such drastic conditions as prolonged loss of sleep must also miss the effects of fatigue when they are used as "fatigue tests" to measure the cost of work. Even where it is possible to show that a given test-decrement is correlated with the duration of work, it probably does not reflect all the fatigue. Until we have measures which indicate the performance of the individual in the more complex activities involving thinking, imagining, comprehension, and emotion, we are bound to miss many of the elements of cost of work when we use the technique of fatigue tests.

It must also be pointed out that tests designed to reflect the effects of sleep-loss reflect changes in capacity rather than changes in efficiency.[41] Not only that, but they may not be completely valid indicators of capacity because of the subject's tendency to leave some "cushion" or "safety factor" even under the high motivation of a laboratory experiment. The lack of sleep may have the effect of reducing capacity on a given test, but at the same time the gap between

---

[40] H. M. Johnson, T. H. Swan, and G. E. Weigand, Sleep, *Psychological Bulletin,* 1926, *23,* 482-503.

[41] See discussion of capacity tests in Chapter 4.

actual performance and capacity is reduced. Performance, therefore, remains apparently little affected by the sleep-loss.

The direct measurement of efficiency of performance following sleep-loss would appear to be a more satisfactory approach to the problem than that which depends upon tests of capacity alone. This is, of course, more easily said than done. For this reason, and also because many investigators do not appear to recognize the need, few studies of the cost of work following sleep are available. Laird and Wheeler have reported a study of oxygen consumption during mental calculation.[42] They reduced the sleep of their subjects from eight to six hours per night for a period of one week. Although the subjects did not decrease their performance (errors remained the same, and speed increased), the oxygen consumption was increased during the "insomnia" period. This study does not have much significance, however, for the following reasons: (1) Although the results were marked and consistent, there were only three subjects. (2) As we have already seen, oxygen consumption is not a reliable indicator of the cost of mental work. (3) Other studies of Laird's showing the effect of mental activity upon oxygen consumption have not been corroborated by other investigators. We mention the study only to show that this general approach has been attempted, and also because the value of the study as an investigation of efficiency is sometimes overestimated in writings upon the subject.

It will be seen that the attack upon the fundamental problems of the nature and effects of sleep has done little more than clarify the problem and show up some of the pitfalls which await investigators in this field. In the light of this situation, it is understandable if an investigator who wants to achieve results in a practical study of sleep problems goes back to more qualitative methods, and to a dependence upon the reports and comments of his subjects. This is essentially what Kleitman has done in an experiment upon the effects of certain foods taken before retiring.[43]

In part of the study, Kleitman recorded movements during sleep as an indicator of depth of sleep, but the bulk of his conclusions was based upon a questionnaire which was answered by the subjects each day. The subject reported upon the duration of sleep, the presence or absence of dreaming, the "quality" of the sleep, whether or not he felt refreshed when he awakened, and upon incidents of the day

---

[42] D. A. Laird and W. L. Wheeler, Jr., What it costs to lose sleep, *Industrial Psychology*, 1926, *1*, 694-696.

[43] N. Kleitman, F. J. Mullin, N. R. Cooperman, and S. Titelbaum, *Sleep Characteristics*, Chicago, University of Chicago Press, 1937.

which may have affected these factors. Various substances, including a popularly advertised "bedtime" food-drink, were administered on experimental nights before retiring. Other nights served as controls. In addition to the effects of the foods taken before sleeping, the effects of weather and season were also analyzed.

All the data showed wide variations from subject to subject, and for the same subject on different nights. Whether this variation is due to the unreliability of the method of obtaining the information, or to valid changes in these variables, is not, of course, known. At any rate, the variation was so great that few significant effects of the controlled variables could be discovered. The authors conclude that the patent food "significantly decreased the motility during sleep," and that the patent food "is the only material which, when taken prior to going to bed, increases the percentage of mornings on which one awakens well rested." [44] In light of the data presented by these authors, these statements would seem incautious, especially since they provide the advertisers with ammunition for a high-powered sales campaign. The levels of significance which were accepted were not very stringent for conclusions which are to be used in this manner, and there is not even internal consistency of the results. It was true that the patent food resulted in significantly more reports of "feeling well rested" when the nights in which this food was used were compared with the *control* nights (when nothing was administered). The effects of the food were not significantly different from the results for nights in which water or milk was given *without* the food. In fact, the effect of cold water upon this index was almost as significant when the comparison was made with the control nights.

There are similar inconsistencies in the results with the objective measure of motility during sleep. The patent food was significantly effective when administered in cold water or hot milk, but not when it was given in hot water or cold milk. In fact the probabilities vary all the way from .90 down to less than .01, depending upon the medium in which the food was administered. In the case of motility, however, there was a moderately significant difference between the group of all nights when the patent food was administered, and the nights when milk or water was administered alone.

The difficulty of drawing significant conclusions from studies based upon such qualitative indices is not only in the extreme variability of results, but also in the fact that we have no way of determining what meaning they have for the general efficiency of the

[44] Kleitman, Mullin, Cooperman, and Titelbaum, *op. cit.*, p. 85.

individual. Of course, if an individual feels refreshed after sleeping, that in itself is a gain. But does this fact mean anything more than that? Is there any relation to his continued efficiency through the day? Until the answers to such questions as this are known, we could not interpret Kleitman's results, even if they were highly consistent and statistically significant.

Chapter 7

WORK METHODS AND EFFICIENCY

The method of performing any given task is probably the most important factor affecting the efficiency of the worker. We shall include under the heading of *method* any intrinsic aspect of the job which may be changed without changing the nature of the product. Thus there may be a choice of tools to be used, a choice of designs for any given tool, variations in the movements or posture of the worker, the order in which various parts of the task are carried out, the rate at which he works, and the load which he carries.

The problem of choosing the particular combination of these factors which will be most effective has been frequently left to accident, tradition, or unguided opinion. In recent decades, and particularly in very recent times, there has been a trend, however, toward a more systematic study of these questions. Motion and Time Study, or Methods Engineering, includes a body of rules and procedures for designing more adequate ways of doing a job, especially the more routine types of work in industry.

Many of the questions which arise in Methods Engineering are strictly engineering problems. Such questions are, for example: "How fast can a given machine be operated?" "Which kind of tool will cut most rapidly and effectively?" "How can the machines in this plant be arranged to minimize the waste space, transportation, and storage of materials?" Unless the answers to these questions affect the performance of the worker, they are not our concern. Consequently our discussion of Methods Engineering will be concerned only with certain phases of the total field.

Along with procedures for answering questions like those just cited, Methods Engineering contains a body of suggestions for the most effective use of the worker, this part of the field usually being known as "motion study." In brief, motion study consists in analyzing the movements to be made by a worker in doing a certain task, deciding upon the best method, trying it out in order to determine whether it produces worth-while improvements in performance, and training the worker to use the new method. In designing

the new method, the stated aim of the engineer is to reduce the effort required to produce each piece, so that production can be increased at no additional cost to the worker.

Much of the early development of these ideas is credited to Gilbreth, who set down many suggestions which are still in use. The engineer first makes a detailed study of how the job is now being done, taking motion pictures, timing the worker with a stop watch, and sometimes using other devices to secure accurate measurement and description. This analysis commonly results in a micromotion chart, or process chart. The analyst then examines each phase of the task critically, in order to determine whether unnecessary time or effort is spent on the phase. In making his recommendations for change he is guided by certain general "principles" which were first drawn up by Gilbreth and have since been modified and expanded slightly. The best way to realize the nature of these principles is to see the list in its entirety, and then to see examples of jobs to which the principles have been applied.

Barnes, writer of a recent textbook in this field, summarizes the rules of motion economy in the following list [1]:

(*a*) Rules for the Use of the Body

1. The two hands should begin as well as complete their therbligs [2] at the same instant.

2. The two hands should not be idle at the same instant except during rest periods.

3. Motions of the arms should be in opposite and symmetrical directions, instead of in the same direction, and should be made simultaneously.

4. Hand motions should be confined to the lowest classification with which it is possible to perform the work satisfactorily.

5. Momentum should be employed to assist the worker whenever possible, and it should be reduced to a minimum if it must be overcome by muscular effort.

6. Continuous curved motions are preferable to straight-line motions involving sudden and sharp changes in direction.

7. Ballistic movements are faster, easier, and more accurate than restricted (fixation) or "controlled" movements.

---

[1] Reprinted by permission from Ralph M. Barnes, *Motion and Time Study*, (2nd Ed.), 1940, New York, John Wiley & Sons, Inc., p. 144.

[2] The term *therblig* designates a single element of motion.

8. Rhythm is essential to the smooth and automatic performance of an operation, and the work should be arranged to permit easy and natural rhythm wherever possible.

(*b*) Rules for Arrangement of Work to Promote the Above Principles

9. Definite and fixed stations should be provided for all tools and materials.

10. Tools, materials, and controls should be located around the work place and as close in front of the worker as possible.

11. Gravity feed bins and containers should be used to deliver the material as close to the point of assembly or use as possible.

12. Drop deliveries should be used wherever possible.

13. Materials and tools should be located to permit the best sequence of therbligs.

14. Provision should be made for adequate conditions for seeing. Good illumination is the first requirement for satisfactory visual perception.

15. The height of the work place and the chair should preferably be so arranged that alternate sitting and standing at work are easily possible.

16. A chair of the proper type and height to permit good posture should be provided for every worker possible.

(*c*) Rules for Design of Tools and Equipment

17. The hands should be relieved of all work that can be performed more advantageously by the feet or other parts of the body.

18. Two or more tools should be combined wherever possible.

19. Tools and materials should be pre-positioned wherever possible.

20. Where each finger performs some specific movement, as in typewriting, the load should be distributed in accordance with the inherent capacities of the fingers.

21. Handles, such as those used on cranks and large screw drivers, should be designed to permit as much of the surface of the hand to come in contact with the handle as possible. This is particularly true when considerable force is exerted in using the handle. For light assembly work, the screw-driver handle should be so shaped that it is smaller at the bottom than at the top.

22. Levers, crossbars, and handwheels should be located in such positions that the operator can manipulate them with the least change in body position and with the greatest mechanical advantage.

All these rules can be roughly summarized in a single sentence: Arrange the work so that the worker can use the smallest number of movements and muscular contractions possible, ensuring that the movements which remain are symmetrical and smooth, with a minimum of fumbling and variation. More specific rules are needed, however, in order to remind the beginner how he is to accomplish these general results. The trainee also needs a considerable amount of observation of these methods in practice, so that he can more readily learn the technique of applying them. For this reason the mere listing of the rules as we have presented them here must be supplemented by films, or, better still, by practical laboratory courses in the practice of motion study.

Criticisms of Motion Study.—Most of the criticisms which have been leveled at the Methods Engineer involve the contention that the methods are *not* more efficient, from the standpoint of *total* efficiency as we use the term in this book. Nevertheless it should be recognized that the *aim* of the Methods Engineer is to increase efficiency. Whether or not the aim is achieved is the main question before us in this chapter.

Let us first examine how the Methods Engineer evaluates the efficiency of a new method. The usual procedure seems to be the following: The engineer takes motion pictures of the present method of doing the job, or he times it with a stop watch. In either case he records the results of only a small number of complete operations ("cycles"). In fact, if his analysis is based upon motion pictures, he may obtain his time records from only one cycle which appears to him to be "typical." There is rarely any statistical analysis of the results to determine their reliability. After the Methods Engineer has devised his new method, a worker is trained to use it, and after some practice the time record is taken again, as before.

Thus the decision about the value of the new method is based upon speed measurements taken under special circumstances, with a small number of records, for only one or for a few subjects. The lack of reliability data is, of course, not an intrinsic characteristic of the method, and the deficiency could be remedied. Also it should be noted that the changes in method frequently produce such large changes in production that reliability analysis is not so essential. The fact that the analysis is based upon speed, and that it is recorded under unusual circumstances, is essential to the procedure as it is now used. In other words, there is no direct measure of the cost of work. It is assumed that the worker is putting forth equal effort during both sets

of observations, and that variations in output must, therefore, reflect variation in efficiency.[3]

Adverse criticisms of the method have been based largely upon this point, but, unfortunately for progress, they have not produced any more definite evidence of the cost of work than that given by the Methods Engineers themselves. Thus Poffenberger[4] points to the shortened cycle of work as increasing the homogeneity of the task, the factor which Robinson and Bills found to be a cause of rapid "fatigue decrement." This is an *a priori* argument without any evidence to show that the blocking and decrement observed by Robinson and Bills are, in fact, increased by the greater homogeneity of the task. Until this question is examined experimentally, one can argue just as easily for the opposite position. For one thing, the normal duration of a cycle in an industrial task is seldom so short as the duration involved in Robinson and Bills' studies. For another, Robinson and Bills also found a factor of conflict to be important in determining decrement. It will be noted that the effect of many of the rules of motion study is to reduce the conflict in the task. One never has to decide which hand to use, which tool to use, or make any of the other decisions that are likely to be involved in unstandardized methods of work.

Another criticism is that a new basis of payment usually accompanies the change in method of doing the work if the work is on a piece-rate basis.[5] The reason for this is readily seen. If the task is now easier than it was before, the worker should not be paid as much per unit of output as he formerly received. An adjustment can be made so that the worker produces more per unit of pay, thus making a saving for his employer, and at the same time the worker can earn a larger total pay than he did under the former system. The objection to this is based upon the belief that the new method may permit faster production at the expense of additional cost to the worker (by affecting his health or his satisfaction). In other words, the new pay rates may induce the worker to exert himself more in order to maintain or increase his earnings, and the additional production may not represent a true increase in individual efficiency.

---

[3] In practice, the engineer may use the method of "leveling" which is discussed in Chapter 10. This procedure is supposed to allow for variations in effort or other conditions, so that it would be possible to compare rates of work and regard them as indices of efficiency. This method, however, is itself open to question, since it depends upon the judgment of the observer and does not involve any validated measures of effort.

[4] A. T. Poffenberger, *Principles of Applied Psychology*, 1942, p. 392.

[5] E. Farmer, Time and motion study, Industrial Fatigue Research Board (Great Britain), 1923, Report No. 14.

This is a very difficult criticism to meet. If the pay rate is left at its original level, and if the work is actually made easier, then the worker could reduce his effort and still earn more than before. As a result, his production would not reflect accurately the benefits derived from the new technique of work. To reduce the piece rates, on the other hand, is to assume that the saving in effort brought about by the new method is known. The Methods Engineer assumes that the time records in his trial studies do reflect the relative ease of performance.

Another possible way out would be to compare the rates of performance in the alternative methods under a system of time payments. If we could be sure that the worker would not be influenced by habits and traditions concerning what is a normal day's production, his effort should be reasonably constant in the two cases. The requirements of a study which would clearly indicate the relative efficiency of the two methods are not easy to fulfil. Each method should be followed over a long period of time—time enough to cancel out the effects of practice and random variations in effort, and time enough to discover any cumulative effects of fatigue. In addition, the records should be obtained under conditions which are as normal as possible; not a few cycles while the time-study man is looking on or taking motion pictures. These requirements are difficult enough, but still more difficult is that of maintaining constant motivation of the worker so as to ensure that he works at a comfortable pace and that he does not aim at some fixed total output regardless of method.

Obviously, such an elaborate experiment could not be carried out each time the method of work is redesigned for each of hundreds of tasks. Perhaps the only feasible solution is to carry out such studies for some sample cases involving a variety of tasks to which the general principles of motion study have been applied. Such a program would at least provide some factual basis for discussing the criticism under discussion. Meanwhile it is another criticism which is debated by both sides on the basis of opinion and uncontrolled observations alone. The debate will continue along these lines until some of the methods of measuring efficiency employed by the methods engineers have been more satisfactorily validated.

A third important criticism of Gilbreth's methods, and one which is still applied to his followers, is that the procedures ignore the role and importance of individual differences among workers. Some people, for example, may find it much easier to work with the two hands simultaneously. A person who finds it difficult to work with two hands might have such a high level of dexterity with one hand

that he could compensate for the advantage others would gain by the two-handed method. Balchin has given an example of a study of three workers.[6] Among them, two were most rapid when using one method, while the third worked better with another method. There is also an indication, from correlational studies of various kinds of dexterity, that these abilities are relatively independent of one another.

It should be mentioned that this phase of the problem was not entirely ignored by Gilbreth. He assumed that the worker would be selected for the task according to his capabilities. While this assumption would be all right in principle, and much has been done toward improving methods of selection, the required accuracy of selection is impossible at the present time. To select in such a way as to overcome the objection to motion-study methods would require that we be able to choose those workers who have high aptitude for a *particular method* of doing a job, as well as choose those who are well suited for the general kind of work. In other words, we should have to choose workers who are dexterous in the types of movement involved in the job, and in the particular combination and co-ordination required by the motion study. This would carry the process of selection far beyond the range of practical feasibility, even if techniques of measurement were available for carrying out such a program.

Perhaps the answer to this objection lies in the current trends in the application of motion-study principles. In order to give these ideas widespread use, many supervisors, and even workers themselves, are being taught the general principles of motion economy. It is hoped that the supervisor or the worker himself will be able to put these suggestions into practice without the services of a specialist in motion study. If this is done, we should expect that individual differences between workers could be more readily taken into account.

The criticisms we have discussed have been those which applied even to the more careful use of motion-study methods. Needless to say, there are many others which would be evident in instances where the techniques have been carelessly or unscrupulously used. If changes in method do not actually make the work any easier, and if, at the same time, they are made the occasion for reducing pay rates on the job, then serious trouble will result. It is rumored that so-called motion studies have sometimes been no more than an excuse

---

[6] N. Balchin, Movement study in packing, *Journal of the National Institute of Industrial Psychology,* 1931, *5,* 274-275. Also in B. V. Moore and G. H. Hartman, *Readings in Industrial Psychology,* New York, D. Appleton-Century Co., Inc., 1931, pp. 298-300.

for reducing pay rates, and as a result the terms "efficiency expert" and "efficiency engineer" have among many workers come to designate a certain disrepute. Sound motion study is not open to this accusation, and those seriously engaged in this work realize how important it is that such suspicions be allayed. The criticisms we have discussed are not to be taken, therefore, as an indictment of the good faith of motion study engineers. They merely raise the question whether the methods have produced as complete a solution of the problem of efficiency as we should have, and whether the manner of gauging efficiency is in serious need of supplementation.

Special Studies of the Effectiveness of Machines and Tools

The rules numbered 20, 21, and 22 in Barnes' list are concerned with the optimal design of certain tools, and are only examples of the general problems which come under this heading. Answers to questions of design of tools have been sought by a variety of methods, ranging from simple time study to the measurement of metabolism.

A very early problem which arose in Taylor's work with heavy labor was that of the choice of size of shovel and load in handling various materials. Taylor designed various shovels which would pick up the proper load of a given material (21½ pounds, according to his analysis). The proper load was arrived at by time study. That is, Taylor chose the load which would permit the highest average rate of performance by the worker. The same problem was attacked later by Wenzig, using metabolic measurements to determine optimal load, height of throw, and distance of throw.[7] He found that the optimal load varied with the height to which the load was to be thrown. For heights over one meter, the optimal load was 7.25 kg. (16 lb.), while for lower heights the optimal load was 9 kg. (20 lb.), the latter a close agreement with Taylor's result.

This example shows the varied procedures which may be used in solving questions of design. One may choose an arrangement which "looks good," or which has grown up by tradition. One may take Taylor's approach and perform time studies with various possible designs, or one may use appropriate measures of the cost of work and so study efficiency directly. The traditional approach is not likely to give the best answer. The choice between the other two depends upon the frequency with which a given tool is to be used, and the availability of measures of cost of work. Shovels are used by so

[7] Summarized in E. Atzler, Arbeitsphysiologie: II, *Ergebnisse der Physiologie,* 1939, *41,* 213-215.

many thousands of workers that it would seem worth while to devote considerable time and expense to determining the basic factors which make for optimal design. Wenzig's study would represent a good start, at least, toward such an analysis. A tool which is designed for a specialized task could not, of course, be studied in this exhaustive manner, and time study would be a more appropriate method. A few examples of tools which would warrant considerable study are the typewriter, automobile and truck controls, screw drivers, and other common hand tools. In addition, the general problem of location and position of levers and cranks would have implications for the design of many machines.

It may not always be necessary to use the slow and cumbersome method of metabolic measurement in order to arrive at a useful answer. In some cases the metabolic measures would not provide a satisfactory answer because the work is too light. Nevertheless it would be well to improve upon time study as a method of answering questions of design for those tools which are in very wide use, such as those just mentioned.

One device which has been applied several times is to compare the force which the worker can exert upon the work with various tools. For example, screw drivers have been compared to determine the shape of handle with which the worker can exert the maximum torque upon the screw.[8] In the latter case, a cylindrical handle was more effective than a tapered handle, and larger diameters were better up to the maximum tested (40 mm.).

Another somewhat similar method has been applied to the typewriter. Norton compared brands of typewriter in terms of the total physical work required for operating the various keys and levers.[9] Among the five brands tested, the worst required about 30 per cent more work than the best. In this case the measurement can be made independently of the operator—it is a measure of the force required to produce a given result. In other problems, as in the case of the screw driver mentioned above, the experiments must be performed with the subject since we are interested in the way the tool fits his hand.

Another problem of fitting the tool to the worker also involves the typewriter. Since the fingers differ in relative strength, the key-

---

[8] B. Rubarth, Untersuchung zur Bestgestaltung von Handheften für Schraubenzieher und ähnliche Werkzeuge, *Industrielle Psychotechnik,* 1928, *5,* 129-142.

[9] F. H. Norton, The work required to operate several makes of typewriter, *Transactions of the American Society of Mechanical Engineers,* 1928, *50,* (Management Section) 29-36.

board should be so arranged that the stronger fingers do more of the work. Hoke, by determining the relative frequency with which different letters are written, was able to show that the present typewriter keyboard is unsatisfactory from this standpoint. As Table 14 shows, the little finger of the left hand has more to do than the second finger on the right hand. As measured by tapping tests, the fingers of the left hand are slower than those of the right, but on our present keyboard the left hand writes more letters than the right. Hoke and others [10] have proposed reorganized keyboards which have received some attention, but so far the difficulties of changing keyboards have outweighed their possible advantages. Of course the arrangement should take into account not only the relative frequency of individual letters, but it should also arrange for common letter combinations to be written with an easy pattern of movements—perhaps alternately with the two hands, as Dvorak and his collaborators have done.

TABLE 14*

LOAD ON FINGERS DURING TYPING

Percentage of Ideal Load	Finger				Average
	1	2	3	4	Load
Right hand	135.8	58.4	100.5	30.6	81.3
Left hand	149.3	153.0	73.1	93.9	117.3

* R. E. Hoke, The improvement of speed and accuracy in typewriting, *Johns Hopkins University Studies in Education,* 1922, No. 7, p. 32.

It should be recognized, however, that these studies of force required to operate a machine, or of force exerted by a tool, are only partial solutions of the problem of optimal design. It is conceivable, for example, that a reduction of the force required to operate a typewriter might not always improve its efficiency. Perhaps a certain amount of pressure upon the keys is better than a mere touch, as in the case of the electric typewriter, or perhaps the pressure required is a very unimportant element in the fatigue of typing. The writer is not a professional typist, but his own experience fails to reveal any particular strain of the hand and arm muscles in typing—the principal feelings of fatigue are postural and are localized in the back. The final test of the design of typewriters, whether from the point of view of pressure, or from the point of view of keyboard arrangement, will not come until it is possible to get accurate measurements of produc-

[10] A. Dvorak, N. I. Merrick, W. L. Dealey, and G. C. Ford, *Typewriting Behavior,* New York, American Book Co., 1936.

tion under comparable conditions for the different designs, and until it is possible to measure the over-all effort and fatigue involved.

A similar argument applies to the studies of screw driver handles and similar devices. The shape of handle which permits the worker to exert the maximal force would be the best handle if the screw driver is to be used only for the exertion of great force. In many uses of the screw driver, however, the maximal force is exerted for only an instant. By far the greater portion of the time is spent in rapid turning against little resistance. This has been recognized, in that a specially designed screw driver for light work has been devised.[11]

Many such problems cannot be completely solved at present, because, as we have seen, measures of effort which are now available are not applicable to the kinds of work involved. In heavy work the situation is more satisfactory, and several applications of metabolic measures have been made. We have already mentioned the metabolic study of shoveling, and in an earlier chapter we discussed Atzler's study of pushing and pulling loads on a cart. An additional example is a study of the use of a wheelbarrow for transporting bricks.

Crowden measured the metabolic cost of wheeling a barrow loaded with bricks under varying conditions of load, rate of walking, and design of barrow.[12] Variations in the number of bricks per load affected not only the weight to be moved, but also the balancing characteristics of the barrow. A load of seventy bricks was found to provide more efficient performance than either a smaller or a heavier load if the optimal arrangement of the load was employed. The optimal arrangement of the load, as might be expected, was that in which the point of balance of the barrow was a position which brought the handles level with the palms of the worker in his upright position. Since a substantial part of the energy consumption was attributable to the lifting of the barrow handles when starting, it was possible to reduce this factor of cost by increasing the length of the legs. By trial, it was possible to use longer legs than those originally provided, without interfering with the use of the barrow.

Evaluation of Progress in Methods-Design.—The time-study analysis which methods engineers use to determine the relative efficiency of methods has contributed much of practical value in industry. In many of the applications, the increased production is so

[11] Barnes, *op. cit.*, 234-235.

[12] G. F. Crowden, The physiological cost of the muscular movements involved in barrow work, Industrial Fatigue Research Board (Great Britain), 1928, Report No. 50.

obviously a result of eliminating unnecessary muscular effort that there is little question of the value of the improved methods. To take an extreme case: If the Methods Engineer finds that a third or even half of the operation can be eliminated completely by a simple rearrangement of the work, we cannot question his statement that efficiency has thereby been improved.

On the negative side, however, certain questions remain unanswered, and certain limitations of present techniques must be recognized. In the less obvious cases, the changes consist of a rearrangement of the pattern of work rather than an evident elimination of parts of the task. Here a sound evaluation of relative efficiency cannot be made solely upon the basis of rate of performance. Until measures of effort are available, there is much room for argument on the relative merits of various methods, or of various designs of tools and equipment. Time study is inadequate because it relies upon extremely crude estimates of effort by the engineer, or upon the assumption that effort remains constant throughout the comparison of the different methods.

It is evident that these techniques of motion economy can be most readily applied to routine manual and mechanical tasks. It must not be assumed, however, that their application is limited to factories. Applications have already been made with success in hotels, laundries, business offices, and even in the home kitchen. A highly variable job, such as preparing a meal, can often be broken down into parts which can be partially standardized. At least those portions which are performed most frequently can be treated according to rules for arranging a work place which were listed on preceding pages.

Where the manual operations are only incidental to the main task—as in studying, writing, scientific work of various kinds, or executive functions—the tools of the trade may be arranged in similar fashion. Extreme disorder may detract from time available for the main task in hand. On the other hand, there is danger of reducing the manual operations too far. It may be better for the desk worker to walk across the room occasionally for a book or to get something from a file. At best, better arrangements will contribute little to the main task except to prevent minor annoyances or delays. In the tasks for which the rules of motion study were first designed, these annoyances assume major proportions and make the difference between economical and uneconomical performance.

Chapter 8

INCENTIVES AND MOTIVES

Most of the popular questions concerning psychology in daily work deal with motives and incentives. Why is a certain individual going into a decline of productivity? What appeal will be effective in raising the output of a certain group of factory workers, students, or soldiers? These are the questions which make up the general topic to be considered here. We might also use that popular word "morale," but it is doubtful if we should gain anything by doing so, since the term is so vague and variable in meaning.

Recall that in our earlier definitions we said that the terms *motive* and *incentive* refer in general to factors which raise or lower the level of *effort* which an individual puts into a task. These factors are usually distinguished from others which we called *governors*. The latter factors affect performance also, but they do so through a change in the *capacity*, or limits, of the organism rather than through a change in effort. The distinction is not always easy to draw in a particular situation, and often the same factor may act both as a governor and as an incentive. For example, hot weather may not only reduce our capacity for heavy exercise, but it probably also reduces our desire to try.

In spite of the difficulties of maintaining this distinction in practice, it is (or should be) the aim of any research study of work to determine which effect is produced by any given factor. The effect upon efficiency of performance will differ, and the way the factor can be controlled, or used in prediction, will also differ. For example, if we do not do so much work in hot weather because weather is a negative incentive, and if it is found that the temperature has no real effect upon capacity, then additional incentives may be just as good as an air-conditioning system. If temperature is actually a governor of performance, however, additional incentives might produce lowered efficiency even though they do raise the level of output.

Incentives and Efficiency.—While an incentive affects the amount which an individual does, it does not necessarily raise efficiency at the same time that it raises productivity. In most physical or muscular

tasks, it is found that the energy consumption per unit of work is least for a moderate pace—a brisk walk rather than a run, and so on. If the effort of an incentive is to increase the pace of the work beyond this optimal level, therefore, it may decrease the over-all efficiency of performance. In the long run, then, it may have adverse effects.

We may assume that any task has its optimal level of effort. For this reason the problem of motivation in work is not simply to find those incentives which will raise output. This is only a part of the program. In addition, we want to find those incentives which will raise production without unduly decreasing efficiency. For example, fear of severe punishment might raise the level of output, but at the same time this fear would waste so much effort that output could not be maintained at the new level without later consequences in the form of ill-health, decline of production, and so on.

There is, then, the need to consider the optimum rate of performance which is to be maintained. The practical use of incentives does not always aim at increasing the rate of production, but rather at maintaining the rate of work at a level as near as possible to the optimum level.

It may be argued that in most industrial operations the level of work is probably below this optimal rate. If so, the problem is usually that of finding incentives which will increase the rate of work. Studies of restriction of output would seem to indicate that this is the normal state of affairs. Nevertheless, upon occasion, there may be the danger of too much effort; the problem then becomes that of reducing incentives to the optimal level.

Sources of Information Upon Incentives and Motives

Results of research on psychological problems, it goes without saying, are always a function of the method by which the data were obtained. In the field of motivation and incentive, however, this point cannot be overstressed. Any particular situation involves so many complex interrelations of motives that any analysis is abstract and incomplete. In our survey of available facts, therefore, we must keep in mind how these facts were obtained. The methods of research study must be noted, along with the facts.

Laboratory Studies of Motivation in Work.—Studying the human being in the laboratory has the advantage of reducing the complications so that it is easier to interpret the results. When we wish to obtain information of practical value on such a problem as motiva-

tion, however, the laboratory situation involves factors which may destroy the very elements we are interested in. If, for example, we wish to find what makes certain kinds of work boring, the fact that our subjects are taking part in an experiment may make them so co-operative that we are unable to produce the type of monotony which would occur in the working situation.

For this reason, laboratory experiments in this field have their main value in preliminary exploration. The results cannot be applied directly to a factory situation without further investigation under more natural conditions. With this caution, let us consider some of the findings which have the most practical value.

In classroom experiments, several of the commonly recognized incentives have been shown to affect speed of learning and performance on various tests. Giving the pupil a statement of his previous performance, without any special effort to incite rivalry, or any praise, produces increased output. Praise, reproof, rivalry, and the presence of other workers have also been shown to be effective. In connection with praise and reproof, further analysis of the effects indicates that praise is more effective for poor performers, and reproof more effective for good performers.

None of these results are very surprising, and they add little more than a confirmation of common sense. Another laboratory study is worth mentioning in greater detail because it is less obvious, and because it has an important bearing upon the setting of standards for piece-rate or other bonus methods of payment.

The experiment to which we refer, carried out by Mace, compared rates of learning with and without a standard or goal set by the experimenter.[1] When the goal was fixed throughout the experiment, the subjects did not progress so well as subjects who were told simply to "do their best." Poorest performance of all was that of a third group who were told to aim at surpassing their previous week's record. (Figure 17.)

We should not jump to the conclusion that setting a goal for performance is poor policy. In fact, Mace draws no such conclusion, but instead tries to find out why these standards were not effective. In the group where a fixed standard (seventy units per session) was given throughout the experiment, he found that the standard *was* influential, but only during the time when the subjects' skill had developed to the point where they could reasonably expect to reach the standard.

[1] C. A. Mace, Incentives: Some experimental studies, Industrial Health Research Board (Great Britain), 1935, Report No. 72.

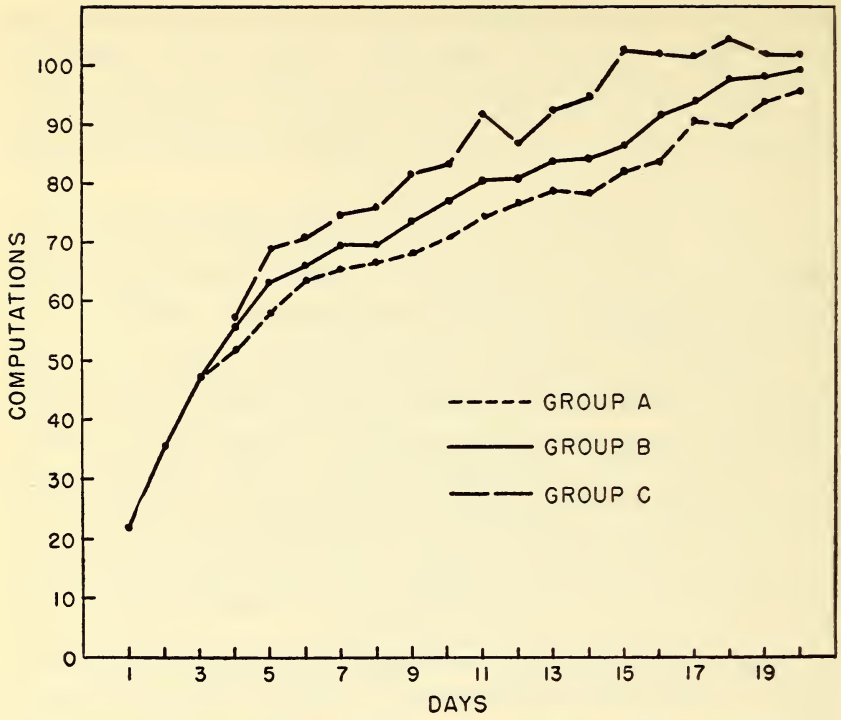


Figure 17. Practice Curves in Computation for Different Instructions

Group A: "Subjects instructed to work with the intention of surpassing a figure representative of their previous performance." (37 subjects)

Group B: "Instructed to endeavor to attain to and surpass a constant and absolute standard, *viz.*, seventy correct computations in the ten-minute period." (37 subjects)

Group C: "Instructed simply 'to do their best.'" (14 subjects)

(From Mace, *op. cit.*, p. 9. Reprinted by permission of the Controller of His Britannic Majesty's Stationery Office.)

In a second part of the experiment, Mace carried the problem further by using a moving standard or goal; that is, each day the subject was told that he was expected to do a certain amount. This amount was estimated on the basis of studies of the rate of learning the task, and was adjusted so that the subject had a reasonable chance of working at the standard rate. Here the setting of a standard produced better results than the instruction to "do your best," the instruction which had produced the most rapid learning in the first part of the experiment.[2] (Figure 18.)

The general conclusion to be drawn seems evident. Setting a goal for the performance of a worker will be most effective if it is ad-

[2] Mace found that the effect of these incentives was to make the daily outupt curve flatter and thus increase the total output without much increase in the maximum rate. If we were to use output criteria of fatigue, it would appear that the incentives increased efficiency as well as rate of work.

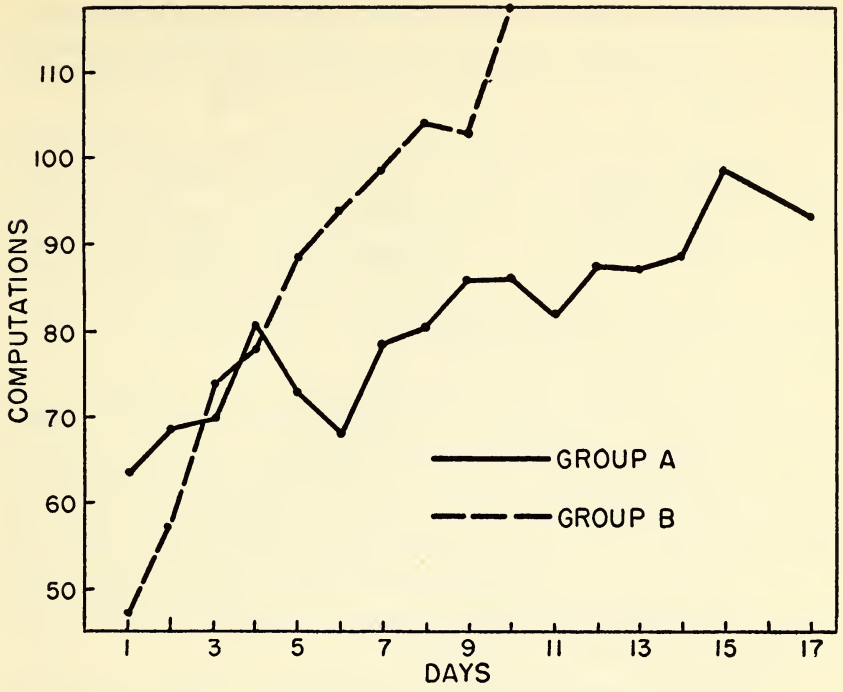


Figure 18. Performance in Computation under Two Instructions

Group A: Subjects instructed to do their best to improve. (No specific standard.) (10 subjects)
Group B: Instructed to surpass a specific standard prescribed for each day. (10 subjects)
(From Mace, *op. cit.*, p. 21.)

justed to his level of skill and ability. Methods of bonus payments which do not take this factor into account may not provide incentives to higher production at all. Some other studies which will be mentioned later would indicate that bonus systems may actually serve as deterrents to production rather than as positive incentives.

Field Studies of Industrial Motivation.—Some investigations are often discussed under the heading of "morale," which is merely another term for the degree of interest and enthusiasm which is felt by the greater number of workers in a particular organization. There are several ways in which such information can be gathered.

1. Questionnaire surveys of attitudes toward company practices and policies, working conditions, and supervision. Here a set of carefully phrased questions is prepared, and provision is made for the workers to hand in their replies anonymously. Much depends upon the wording of the questions, and upon the way in which the

questionnaire is "sold" to the workers. That is, the workers must be convinced that their answers are entirely confidential, and the possibility of distorting the replies through leading questions must be removed.

2. Interview surveys. Here the problems of gaining frank and accurate information are similar to those of the written questionnaire. The advantage of the interview method is in allowing for greater elasticity of the topics considered and the types of comments made. There are corresponding disadvantages, however. The interviewer's bias and poorly phrased questions may invalidate his findings unless he is very carefully trained.

3. Exit interviews. The worker who is leaving because he has been discharged, or because he is disgruntled, is likely to be willing to make statements which he would not make so long as he was employed. Allowance must be made for distortions due to the emotional upset which may have occurred. Even so, the information is of great value, and may permit correction of conditions which could not be discovered by the management in any other manner. Such interviews bring a distorted sample, however, and tell us little of the reactions of the majority of workers.

4. Analysis of turnover records. The turnover rate represents an over-all picture of the effects of personnel policies, supervision, and conditions of work, in relation to economic conditions, competition for labor, etc. If the rate is computed on the basis of *avoidable* losses of men, it gives an even better picture of the effects of company policies and practices. The formula for this computation would be:

$$\text{Turnover rate per month} = \frac{\text{Avoidable separations}}{(\text{Average payroll per period})}$$

Avoidable separations would include dismissals for incompetence and infractions of rules, voluntary quitting for reasons other than health, marriage, etc. It would not include layoffs for reason of shut down in a particular department, or other separations of this kind. The rate thus computed reflects, therefore, the value of the hiring procedure, the relation of wages and conditions to those in other industries and plants, and the conditions of work and supervision.

If the hiring procedure is adequate, and if the wages are standard for the area and industry, then an abnormal turnover rate for a given department, or for a given plant, indicates some source of dissatisfaction which is worth searching for.

Further analysis by department, by supervisors, by various types of worker may aid in the discovery of this factor. Scott, Clothier, Mathewson, and Spriegel,[3] for example, show an analysis of turnover rates for workers of various degrees of intelligence in certain organizations. In one organization the highest turnover was for the workers of higher intelligence, and there was reason to believe that these were more valuable workers on the job involved. Their conclusion was that the supervision had not been adapted to the needs of this group of employees.

5. Naturalistic studies. Interesting hints upon industrial motivation have been gained by men who have simply gone to work in various plants, gained the confidence of fellow workers, listened to conversation, and asked questions. Such reports cannot be considered as the last word on the subject, since too much depends upon the individual bias of the observer, the chance experiences which he has, and other uncontrolled factors. Nevertheless the observations are valuable even if they are considered only as a first step to be followed up by more accurate methods later.

Mathewson's study of restriction of output in industry is an example of the naturalistic type of study at its best.[4] He reports a series of cases in which workers were producing at a level well below what they might have done without overexerting themselves. The important factors which he regards as responsible for this restriction are (1) the desire of the supervisor to maintain as large a staff as possible in order to meet future emergencies; (2) wage "incentive" methods like the piece rate, plus the fear that the rates will be cut if the worker earns too much; (3) the report that a time study is to be taken, as well as the actual presence of a time-study man; (4) fear of unemployment, that is, the desire to make the work last as long as possible.

The general impression one gets from reading Mathewson is that many wage payment plans have been set up by someone who is thinking in terms of "The Economic Man," someone who assumes that workers will always do what is to their greatest *immediate* financial advantage. In other words, the payment methods and other policies seem to be set up without much understanding of the worker. Not only that, but management is often unaware that the payment method is unsuccessful, even after it has been put into practice. Mathewson

---

[3] W. D. Scott, R. C. Clothier, S. B. Mathewson, and W. R. Spriegel, *Personnel Management* (3rd Ed.), New York, McGraw-Hill Book Co., Inc., 1941, p. 502.

[4] S. B. Mathewson, *Restriction of Output Among Unorganized Workers,* New York, Viking Press, Inc., 1931.

gives instances in which he interviewed the managers of a plant and found them confident that restriction of output was a minor problem in their plant because of the new method of payment. In some instances these were the same plants in which Mathewson had observed definite restrictive practices.

Mathewson is aware of the limitations of this kind of study. He notes, for example, that he has no way of estimating the prevalence of restrictive practices. He vouches for the cases which he reports, but he cannot say how typical they are. For such information it is necessary to turn to other sources.

The first two of these methods of field study are better adapted to discover facts on motivation which would have wide application. Methods (3) and (4) are used primarily to find specific sore spots in an organization, information of great importance to the organization concerned but not of as great value to us in our search for general information on the important incentives for industrial workers. At least they have not produced so much up to the present time. Method (5), as we have suggested, is not a very reliable procedure. Therefore we shall devote the major discussion to questionnaire and interview studies.

General Results of Questionnaire and Interview Studies

A number of these investigations have asked the worker to rate, or rank, the relative importance of various factors in the work situation as incentives for remaining at the job or as factors which would determine his changing to a new job. This is a very difficult question for anyone to answer, but it was hoped that careful phrasing of the question, and securing answers from large numbers of workers, would cancel out some of the possible errors. Whether or not this is true, the results are interesting as indicating what the workers *think of* as important influences, even though there may be other less concrete factors which are not mentioned.

Goodwin Watson has summarized several of these sets of data,[5] and the following are reproduced from this source.

Table 15, below, shows the ranks which were assigned to various factors in a poll taken by Fosdick. The first column reports the relative standing of the items in the voting by a large number of employees, while the second column reports the employers' opinions as to the relative importance of the factors as incentives. Thus the

[5] Goodwin Watson, "Work Satisfaction," Chapter 6 in *Industrial Conflict*, Yearbook of the Society for the Psychological Study of Social Issues, 1939, pp. 114-124.

study shows something of employers' misunderstanding of motivation; it also gives us useful hints as to which factors the workers regard as important. Of course we cannot say how other factors might have been placed in relation to the eight mentioned, since these eight were the only ones listed in the question.

TABLE 15 *

RANK ASSIGNED VARIOUS FACTORS IN MORALE BY EMPLOYERS AND EMPLOYEES

Morale Item	Employee Ranking	Employer Ranking
Credit for all work done	1	7
Interesting work	2	3
Fair pay	3	1
Understanding and appreciation	4	5
Counsel on personal problems	5	8
Promotion on merit	6	4
Good physical working conditions	7	6
Job security	8	2

* Watson, *op. cit.*, p. 119.

In Table 16 Watson shows the results of a study of "225 young men, all American born, unmarried, high-school graduates, who had had two or more full-time jobs for at least a year, but who were unemployed in 1933-34 .... They were asked, 'Which of the jobs you have held appealed to you most or proved most interesting?' and 'Why?' " The reasons mentioned most frequently are given in the table. One hundred fifty-seven out of the 225 gave classifiable answers.

TABLE 16 †

REASONS ASSIGNED BY 157 YOUNG MEN, HIGH-SCHOOL GRADUATES, FOR PREFERRING ONE JOB RATHER THAN ANOTHER

Reason	Per Cent
(1) In line with vocational aspiration	29
(2) Congenial contacts with people	24
(3) Like responsibility, chance to use initiative, prestige	19
(4) Variety	12
(5) Opportunity for promotion	8
(6) Salary better	4
(7) Shorter hours	4
	100

† Watson, *op. cit.*, p. 120.

The general pattern which emerges from these data should be kept in mind by every supervisor and executive. Wages turn out to be a secondary consideration, and we shall present other evidence on this point a little later. In general, the worker seems to want a certain degree of independence and initiative, plus recognition for his work and his value to the organization. He wants a supervisor who guides and directs rather than commands.

The failure or disappointment involved in many "welfare" schemes in industry has probably resulted from a failure to take these factors into account. A fancy clubhouse, or a much-publicized profit-sharing scheme may be unsuccessful because it neglects or even prevents attainment of some of the above conditions. The clubhouse does not make up for irritating supervision, and it may appear as charity rather than as recognition of the value of the worker's performance.

The greatest value of these questionnaire studies of motivation or attitude is probably not to be found in these very broad questions which try to get a composite picture of factors for a variety of jobs and conditions. Instead, the greatest benefits come from information which is so specific that it does not lend itself to a general summary of this kind. A particular source of friction in Department A of Company B may not shed much light upon the general problem of motivation, but it may be extremely important to the efficiency of that organization. Perhaps from the point of view of benefits to efficiency, the discovery of this sort of specific information is more important than the attempt to discover generalizations and principles. If that is so, it is more important that the applied psychologist be trained in the *techniques* of developing and administering reliable questionnaires than that he have a consistent and all-embracing theory of the motives of the worker.

The Western Electric Studies.—The Western Electric Company, in collaboration with the Harvard Business School, has been carrying on extensive investigations related to the topic of this chapter. In fact they are the most elaborate studies of the kind ever carried out; it would be impossible for us to do full justice to them here. Roethlisberger and Dickson have published the most recent summary of the results in a large volume.[6] This was summarized by Stuart Chase in "The Reader's Digest." [7]

---

[6] F. J. Roethlisberger and W. J. Dickson, *Management and the Worker*, 1939, Harvard University Press. Also summarized in the National Research Council Report, *Fatigue of Workers*, 1941.

[7] *The Reader's Digest*, (February) 1941.

In brief outline, they started to investigate the effects of such factors as lighting and rest periods upon production. Their procedure was to set up a special experimental room within the factory to which certain operators were transferred to do their usual work. After many months' work with various combinations of rest and work, they found that production was markedly improved. (Figure 19.) When they returned to the original conditions with no rest periods, however, they found that production remained at a high level rather than returning to the original value.

A search for the reasons led them to the conclusion that the important factor in the improved production was not the rest periods. Instead, whatever effect the rest periods had was covered up by a much more complicated set of factors. The operators had been given special attention; they were allowed more freedom than they had in the regular department; the supervisor was more of an observer than the type of supervisor they were used to; and the operators themselves had become a more congenial and closely knit group.

The results led to an extensive interview study of several thousand workers in the plant. Special interviewers were trained to gain the confidence of workers and convince them that what they said would be kept anonymous. Any specific complaints were referred for investigation and correction if possible. Comments about the supervisors were brought up for discussion in supervisory training classes with the identifying information concealed, so that the supervisors could see what effect some of their acts and policies had produced.

In addition to the benefits produced in these ways, the investigators believe that another result was just as important. The workers felt better for having been interviewed in this manner, even if there was no evidence that their statements had an effect. They felt that an interest was taken in their feelings, and the interview gave them a chance to "blow off steam" on topics which they otherwise could not mention and which they might brood over indefinitely.

A substantial part of the value of the published reports we have mentioned cannot be reproduced in a brief space. Interesting observations on social organization within departments are presented. Several of the interviews with workers are reproduced in detail, and reading them gives an insight into the method and the results which can be gained in no other way. The reader is therefore urged to turn directly to Roethlisberger and Dickson's report.

Evaluation of the Western Electric Studies.—Since the "Hawthorne experiments" (as these studies are often called) have received

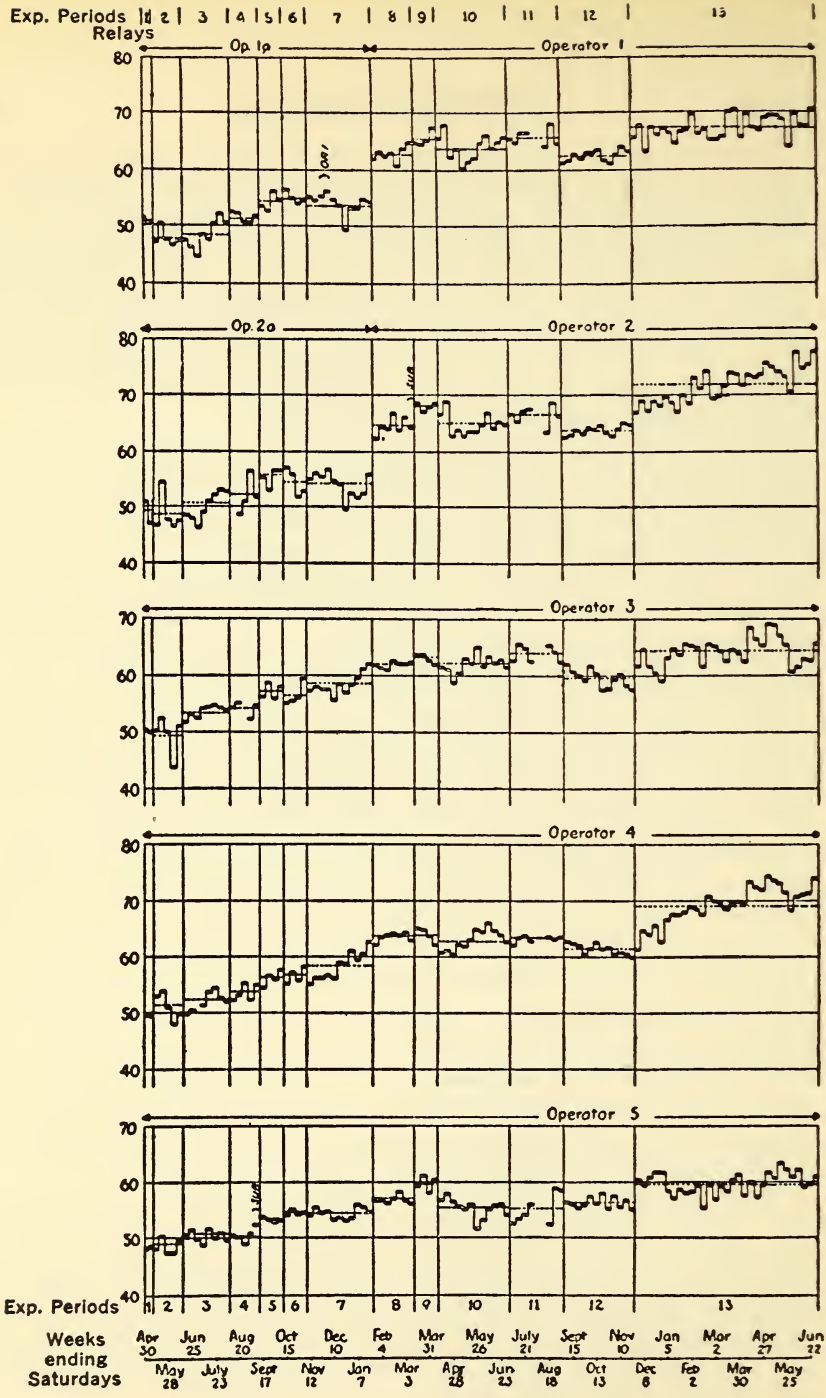


Figure 19. Average Hourly Output in Western Electric Company Test Room Studies

(See table on following page.)

so much public attention, we may well consider just what their psychological importance is. In the first place it might be said that their most important findings were obtained because they began with a faulty experimental method. It is not surprising that any effects of rest periods were covered by more complicated motivational factors when we note that the subjects of the experiment knew that an experiment was going on and that rest periods were the subject of investigation, and when we recall that those participating in the experiment were given special privileges and attention. If we wanted to find the effects of rest periods *per se,* these would be disturbing factors which have to be controlled. Before the value of rest periods could be assessed, it would be necessary to introduce the rest periods with as little change from usual working conditions as possible. Also, the conditions would have to be maintained for a long time so that the effect of novelty, and the feeling that management was taking greater than normal interest in the worker's welfare, would wear off.

We have said that the results were not surprising. They are not surprising now, but the more important question is whether they should have been anticipated before the experiments were started. We cannot answer this now, although it seems that it should at least have been suggested at the time.

It may be true that management should have been aware that workers who feel that their work is important, who are given a chance to express their preferences and opinions, and who feel themselves free from overly strict supervision would work more effectively. Those who planned these experiments probably would have agreed,

Data Pertaining to Figure 19

Period	Length in Weeks	Experimental Conditions of Work	Experimental Working Hours per Week *	Per Cent Decrease from Standard Working Hours
I	2	Standard	48	..
II	5	Standard	48	..
III	8	Standard	48	..
IV	5	Two 5-min. rests	47:05	1.9
V	4	Two 10-min. rests	46:10	3.8
VI	4	Six 5-min. rests	45:15	5.7
VII	11	15-min. A.M. rest and lunch 10-min. P.M. rest	45:40	4.8
VIII	7	Same as VII, but 4:30 stop	43:10	10.0
IX	4	Same as VII, but 4:00 stop	40:40	15.3
X	12	Same as VII	45:40	4.8
XI	9	Same as VII, but Sat. A.M. off	41:40	13.2
XII	12	Standard	48	..
XIII	31	Same as VII	45:40	4.8

* By "experimental working hours" is meant the total time lapse between official starting and stopping time for the day (standard working hours), from which those time decreases due to the experimental conditions of work listed above have been deducted.

(Reprinted by permission of the publishers, from Roethlisberger and Dickson, *op. cit.,* p. 76.)

in principle, to such a statement at the time. They apparently did not, however, realize that their experimental procedure was going to bring about these conditions.

Perhaps more important than any actual contributions to our fundamental understanding of industrial motivation is the fact that the research indicates a trend on the part of management to become interested in these problems and to do something about investigating them. In addition, the wide interest which these studies has aroused may pave the way for further and even more enlightening studies.

The Western Electric Company itself felt that it had benefited in its supervisory relations as a result of the test room studies and the elaborate interviewing programs which grew out of them. What had been an abstract statement became more concrete and intimately related to their work and experience. When we try to draw general statements or principles from these studies, however, we go back to very general statements which seem self-evident.

Perhaps industrial psychological research has other values for the organization which carries it on, beyond the factual and generally applicable information which is obtained. Certainly it studies a situation which may not be very closely duplicated anywhere else, and much less extrapolation is necessary. In addition, it may help to create a greater interest in psychological factors in work among the staff of the organization which carries on the research.

Summary of the Main Industrial Incentives.—In order to bring some order into our collection of information upon incentives in work, it will be well to list some main headings under which these facts may be grouped. The following list appears to include most of the important factors:

1. Pay.
2. Standards and goals.
3. Intrinsic interest in the task.
4. The social setting of the job.
   (*a*) Supervisory relations.
   (*b*) Relations to other workers (competition, conforming to customs, etc.).
   (*c*) Attitudes toward organization or management as a whole.

We have some information which fits under each of these headings, although much more is to be desired. In some of these areas, little more than a start has now been made upon a research approach to the problems. In some cases a single research has been concerned

with a mixture of factors falling under different headings. In such cases it is not always possible to separate our discussion of the various headings, and there will be much overlapping. So far as we can, however, we shall summarize the available results in this fashion.

1. Pay. No one doubts the importance of pay as a determining factor in the choice of jobs, or productivity in a given job. There has, however, been a tendency to overstress this factor in relation to others. The studies we have already discussed indicate that this factor cannot substitute for many other incentives which are necessary to effective work.

"Incentive pay" systems have sometimes tended to overemphasize the importance of pay, and in addition they have assumed that the worker will be influenced by immediate returns. Studies of restrictions of output indicate that where pay is an important factor, the expectation of long-term average earnings is more important than any immediate return which might be temporarily gained by extra effort at the moment. This is one way of summarizing Mathewson's finding that workers are unwilling to work near capacity if it means high earnings for a time, followed by a reduction in rate. In addition, the worker is unwilling to earn more at the expense of disapproval of other workers.

2. Standards and Goals. Mace's study indicates the importance of setting standards for the work which are adjusted to the present ability of the worker. Our discussion in Chapter 10 of the methods of time study now in use in industry for determining standards of performance, as well as Mathewson's study of restriction of output, indicate that this aim is by no means reached by current industrial methods.

3. Intrinsic Interest in the Task. This factor is usually discussed in industry in terms of the opposite of interest—monotony, dislike of the work, dissatisfaction with the nature of the work. Primary determiners of interest or lack of interest appear to be matters of individual differences. A number of researches have been concerned with the discovery of the individual characteristics which make for more satisfactory adjustment to work on a routine, repetitive task, and these will be discussed in the next chapter.

Aside from these matters of individual variation in interest and boredom, a few researches have analyzed factors which decrease boredom for most individuals—general factors of the kind which are the main concern of the first portion of this book.

Thus it has been found that performance on routine tasks can be increased, and the reports of the workers upon their own boredom improved, by certain revisions of the working procedure. Piece-rate payments seem preferable to time payments from this point of view. It is reported better to supply the materials in lots, and perhaps to allow the workers to stop work in order to secure a new supply, rather than keep them continuously supplied with material. In some cases it is possible to allow workers to alternate jobs. This is helpful if the time spent on each job is adjusted to the proper period. If it is either too long or too short, no benefits may be gained. It appears that the change of occupation need not be very great. Perhaps it helps to change the product slightly, even though the general nature of the work is the same.

The improvements resulting from changes of the sort just described are modest at best, but they are probably worth trial for some of the most repetitious tasks.

The dissatisfaction due to the nature of the work is difficult to separate from that due to a variety of other aspects of the working situation, such as the supervision, company labor policies, and the like. Not only that, but the attitude of interest toward the work is a function of these other phases. Not only may the worker dislike the supervisor, but he may also dislike the work he is doing because of his relations to the supervisor rather than because of the nature of the task itself. The improved relations to supervision found in the Western Electric Company test room studies can be regarded as an instance of the opposite effect. In many cases it is possible that the nature and arrangement of the work itself are relatively minor factors in determining contentment and satisfaction. The worker whose task is simple and not very taxing, and whose specific ambitions do not go much beyond his present status, may find that the work is neutral in interest. Much more of his attention is occupied with those factors which we have grouped together under the next topic, the social factors in the work setting.

Our knowledge of the psychology of interest is extremely limited at the present time. We have some research directed at the problems of monotony and dissatisfaction, as already mentioned and as amplified in the next chapter. We also have a considerable body of information about the expressed likes and dislikes of an individual as predictive of his vocational adjustment, and something about the stability of patterns of specific likes and dislikes. We still have far to go, however, in discovering the fundamental nature of interest, its place in working activities, its effect upon efficiency, and the

factors which modify and affect it. So far, we do not even have techniques for studying these problems, or formulations which would make it clear what our research aims should be. Most of our psychological texts which touch upon the topic contain a common-sense discussion which might almost as well have been written by a layman as by a trained psychologist.

While we know little of the nature and conditions of interest, there can be little doubt of its importance in efficiency. Common observation indicates that work of an interesting character is not only more rewarding, but also less costly to the individual. It may be that greater improvements in efficiency can be obtained through increase in interest than in any other manner. There is even a suggestion in one of Poffenberger's studies to indicate that the energy cost of work is reduced by interest.[8] Measuring metabolic rate during mental calculation, the investigators concluded that the output increased and energy requirements diminished when the subject became absorbed in the task as an end in itself. When a monetary bonus was the primary goal, output increased, but the energy consumption also increased. As we have already pointed out, energy consumption is not an adequate indicator of efficiency in this kind of work because the changes are small and subject to the influence of many other factors. Nevertheless Poffenberger's conclusions furnish us with a statement of a problem which is worth much intensive investigation.

In industry, special bonus methods of payment are more widely applied to those tasks which are regarded as intrinsically boring and distasteful. These bonuses, which are supposed to substitute for interest in the task itself, may reduce efficiency even if they are successful in raising output. Another hint of this same relationship is to be found in Wyatt's experimental study of different methods of payment.[9] An experimental group of girls worked in rotation upon five different tasks, spending one day per week upon each of the five. Three systems of payment—time-rate, bonus, and piece-rate—were employed at different times. The effects of these incentives varied from worker to worker, and also from job to job. In one task, for example, the output under piece rate was nearly three times that under a time rate, while another task was increased by only one half. (Figure 20.) The order of effectiveness of the wage incentive for

[8] From an unpublished study by A. T. Poffenberger and G. H. Rounds described in A. T. Poffenberger's *Principles of Applied Psychology*, 1942, pp. 410-411.

[9] S. Wyatt, Incentives in repetitive work, Industrial Health Research Board (Great Britain) 1934, Report No. 69.

different jobs was almost the same as the order of preference expressed by the workers in rating the jobs.

Although this experiment of Wyatt's represents only a beginning in research along this line, it indicates again the possibility that piece rates or other bonuses may be applied to the kinds of work in which they are least effective because of the low level of interest involved

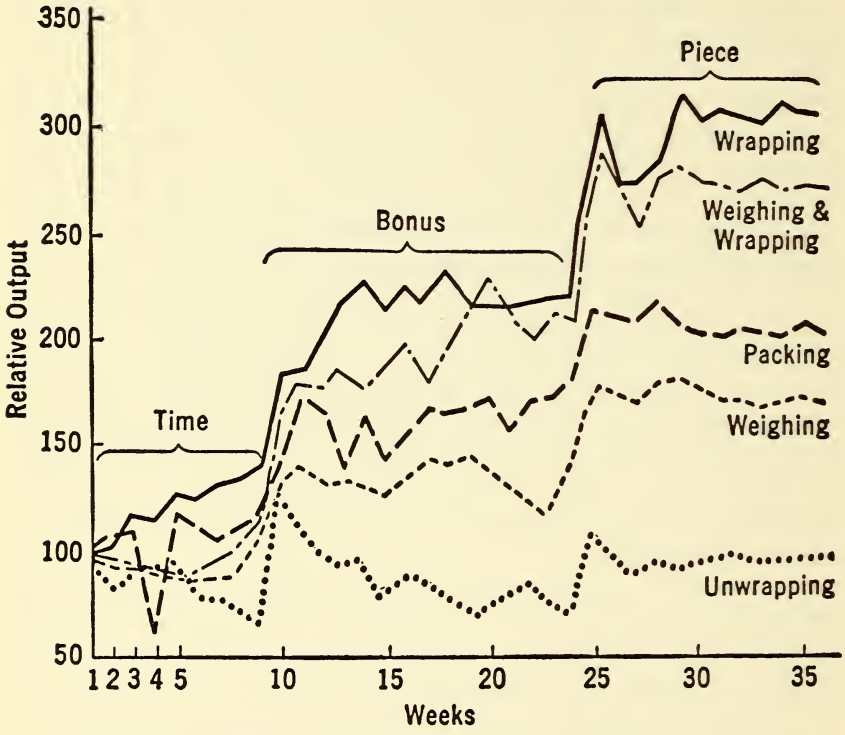


Figure 20. Effects of Incentives upon Different Kinds of Work

(From Wyatt, *op. cit.*, 1934, p. 24. Reprinted by permission of the Controller of His Britannic Majesty's Stationery Office.)

in the job. That is, we are suggesting the possibility that extrinsic incentives like pay, in the absence of the interest factor, increase output at the expense of increased cost of work, and also that they are not even very effective in increasing output.

4. The Social Setting of the Job. While the importance of supervisory relations, interchange with other workers, and the attitudes of the worker toward management are factors of recognized importance in effective performance of a job, experimental investigations like those of the Western Electric Company have done little

more than show how important these factors can be. If we look for specific recommendations on how these factors can be controlled and made the most of in practical affairs, we find little more than a statement that we should be attentive to these phases of the working situation, and that we should use good judgment in dealing with problems of this kind.

Although the recommendations for practice are broad and vague, we have made some progress in developing tools for fact-finding in this area of motivation. By the use of interview techniques, questionnaires, or others of the techniques discussed in this chapter, it is possible to get a more accurate estimate of the workers' attitudes toward company policies, adequacy of supervisory methods, specific sources of irritation and discontent in the organization, and physical layout of the plant. With this information available to supervisors and policymakers they can use their judgment and common sense on the basis of facts rather than on the basis of inaccurate guesses and hunches about worker reactions to policies.

Because of the vast number of factors which enter into these social relationships of the worker, it is doubtful if recommendations for practice can ever include specific supervisory procedures or managerial policies, and at the same time be generally applicable. If a given supervisor or manager can, however, be provided with reasonably accurate and representative data upon the specific things which are causing discontent in the organization, he is already far on the road to a solution. In other words, it is likely that in the past errors have been due as much to inaccurate information about the current situation as they were to faulty notions about what to do about the facts that were available. The manager who is surprised because he has labor trouble in spite of the fact that he has provided elaborate "benefits" for his workers—bowling alleys, clubrooms, health insurance, etc.—is surprised because he has no knowledge of how these benefits are regarded by the majority of the workers, and because of other factors which are considered more important by the workers themselves. It is not that he has violated some general principle of motivation because these same benefits may be very effective in some other organization where the conditions are somewhat different.

Chapter 9

"NERVOUS" FATIGUE AND BOREDOM

It is commonly recognized that the tiredness of a worker depends in a very intimate way upon "emotional" and "motivational" factors, and that these factors are frequently more important than the effect of the work itself. The man who is continually irritated by his working mate or his superior, who is in fear of discharge or of serious accident, who feels that he is doing unimportant work which stamps him as a failure, or who actively dislikes the tasks he is required to perform—such a man is likely to be exceedingly weary at the end of the working day, and indeed may feel "worn out" most of the time.

In view of the great importance of factors like those we have just described, coupled with the difficulties of measuring effort and fatigue which we have previously discussed, there has been a tendency on the part of some writers to pay little attention to the nature of the work and its effects upon the organism, unless the work requires severe muscular exertions or is performed under extreme conditions of heat, pressure, oxygen tension, or the like. The National Research Council Committee which published "Fatigue of Workers," [1] for example, devoted the greatest portion of that report to a discussion of the Western Electric Company studies which are described in the preceding chapter.

Because of this trend to emphasize the "attitudinal" and "emotional" side of the working situation, we must attempt to relate these factors to the general discussion of efficiency and the cost of work previously presented. It will be necessary to proceed on a commonsense basis without adequate support from factual studies because that is the present status of this problem.

"Nervousness," emotive predicaments, boredom, and distaste for the work contribute to decreased efficiency in a multiple fashion. First, the useful accomplishment of the worker in terms of "objective" output is partly or wholly canceled by the dissatisfaction which

[1] Committee on Work in Industry, National Research Council, *Fatigue of Workers: Its Relation to Industrial Production,* New York, 1941.

he feels. In other words, the net value of the accomplishment is low if we count as output not only the pieces produced but also the satisfaction which the worker derives from that production. Second, effort is reduced, and consequently the objective output is also diminished. Third, while the effort directed toward the work is reduced, the emotion involved in the day's activities costs the organism a great deal. At least the feelings of tiredness are markedly affected in our common experience, and we should expect that these feelings must reflect some change in total bodily economy.

While a common-sense analysis of the influence of these factors upon efficiency is easy to carry out, a more scientific study is extremely difficult. In fact, scientific study of the problem has been almost nonexistent. As we have seen, we have the beginnings of an analysis of the factors which produce dissatisfaction and irritation in industrial work. One does not have to look far to find very potent sources of irritation. What we do not know is the effect of this irritation upon the organism. At this point it may be well to consider the reasons for this serious gap in our knowledge. There are several important difficulties involved.

The degree of upset or irritation, for example, must be assessed by questioning the workers. The unreliability of such a procedure is well known. As a result, many studies which have touched upon the problem have resorted to analyses of production rather than of the irritation itself. In the Western Electric Company studies they drew their conclusions about the improved social setting of the job partly from casual comments and remarks of the workers, but primarily from the fact of continuously increasing output which could not be accounted for in terms of greater ease of the work. Though we do not doubt the general validity of the conclusions of this study, it should be noted that it is based upon rather unreliable criteria. As it happened, the experimental conditions produced such marked changes in the feeling tone of the workers that there was little danger of error due to the inaccuracy of the methods. In other problems dealing with less drastic changes in the working conditions this unreliability becomes a serious difficulty.

Besides the difficulties of determining when irritation and dissatisfaction occur in a given working situation, and also how much, there is the further problem of determining their effect upon the individual. Aside from the low production rate, for example, and the fact that the employer may lose a trained worker, does it do any harm if that worker is chronically bored with his task? Production rate and turnover are, of course, important in themselves. It is also possible that

chronic boredom which cannot be escaped may lead to deterioration in the worker. The effects, however, are likely to be extremely complex; they cannot be discovered by physiological measurements of the worker before and after a day's work. There are likely to be progressive changes in his general adjustment at home and in his hours of recreation. Alcoholism, divorce, or petty fights and disputes may have some basis in the worker's dissatisfaction with his job. Occupational neuroses may develop.[2] While these relationships appear obvious, nevertheless it is still impossible to determine the exact nature and frequency of the dependence.

We shall presently examine the researches which have been devoted to the problem of boredom in work. Before doing so we must discuss the place of studies of fatigue and effort in relation to these problems.

First let us recall that the cost of work is conceived in this book as including *all* adverse effects of the work upon the worker. Thus, at least in principle, we have provided a place for "nervous" fatigue and monotony in our general formula for efficiency. Moreover, some of the measures of cost of work which have been described and analyzed make no distinction as to the origin of the effect—whether in the working activity itself or in its surrounding conditions. Circulatory changes such as pulse rate and blood pressure, and changes in muscular tension will obviously be a function of emotive factors in the work as well as of the effort required by the task. In the laboratory studies in which these measures have been used, the social and emotional factors have been held at a fairly constant level. The pattern of changes which comes about in work with disagreeable accompaniments in the surrounding situation will require investigation.

We are handicapped in our analysis of this problem by the fact that most experimental work upon the bodily changes in emotion has concerned temporary emotional upsets rather than the long-drawn-out but milder anxieties, fears, and angers frequently involved in the normal working situation. Our interest is in the repetition of emotional seizures, or the chronic emotional strain due to conflicts with the supervisor, distasteful work, resentment against management policies, and the like. Whether this kind of upset has organic effects similar to those which occur in the sudden startle or fear of the laboratory situation is difficult to determine. It is quite likely that there are permanent functional-organic changes—circulatory, vis-

---

[2] See M. Smith, M. Culpin, and E. Farmer, A study of telegrapher's cramp, Industrial Health Research Board (Great Britain), 1927, Report No. 43, and M. Smith and M. Culpin, The nervous temperament, *ibid.*, 1930, Report No. 61.

ceral, and glandular—which result from these chronic emotional upsets, but so far we have little experimental evidence to indicate their nature. The reason for our lack is obvious. We cannot set up such situations on a deliberate experimental basis. To follow the effects of the natural working situation involves so many complex and uncontrolled variables that it is unmanageable.

Two lines of indirect evidence may eventually lead to techniques applicable to the working situation. Physicians in the field of psychosomatic medicine have recently aroused an interest in the emotional factors contributing to organic disease (asthma, heart disfunction, digestive disturbances, and so on).[3] The disturbing factors most frequently cited involve the family, marital, and general social relationships of the individual rather than his job situation. Nevertheless it may later be possible to assess the importance of emotional factors in work as contributing to psychosomatic disorders.

The other line of indirect evidence comes from studies of animal behavior. It is possible to develop "experimental neuroses" in lower animals, and to study their visceral and circulatory functions both in the experimental situation and outside.[4] While these studies are making important strides toward the achievement of a better understanding of many aspects of neurosis, it is too early to attempt the dangerous step of applying the results to the complex conditions of shop and office.[5]

The gross statistical indices of cost of work in industry also represent composite effects of all the conditions of work. In fact, turnover rates and absenteeism in industry may often reflect the emotional factors more clearly than they do the effort required by the task. Similar relationships probably hold for health records, although the factors in this case are more complex. These gross measures cannot, of course, reflect individual reactions to the working situation. It is only when a substantial proportion of the workers are affected by

---

[3] H. F. Dunbar, *Emotions and Bodily Changes* (2nd Ed.), New York, Columbia University Press, 1938.

[4] H. S. Liddell, Conditioned Reflex Method and Experimental Neurosis, Chapter 12 in *Personality and the Behavior Disorders,* J. McV. Hunt, Editor, New York, The Ronald Press Co., 1944, pp. 389-412.

[5] There are evident similarities between some of the situations which produce animal neuroses and those which occur in jobs where the individual must remain at work performing a task which he dislikes or cannot perform adequately. There are probably similarities in the effects of such situations in lower animals and in man, but we do not know how close the similarity may be. Liddell, for example, has experimentally produced in goats a "stiff leg" reminiscent of telegrapher's or writer's cramp, as well as conditions of "anxiety." The factors which produce these conditions in animals are certainly not identical with those which produce similar symptoms in man.

poor supervision, unfair management policies, or distasteful tasks that these conditions are reflected in turnover and absentee rates. Variation from individual to individual in the effects of these factors must be studied by more refined methods.

In the Western Electric Company studies there is no direct evidence on the cost of work in the test-room experiments. Inadequate measures of fatigue were tried out with negative results (output curves for the hours of the day and the days of the week, vascular skin tests, blood pressure tests, and speed tests on the job). From these results they conclude: "There is no evidence in support of the hypothesis that the increased output rate of all these operators during the first thirteen experimental periods was due to relief from fatigue. This hypothesis, which at first seemed most plausible, had to be abandoned." [6]

We may accept the conclusion that the primary reason for the increases in output in the Hawthorne studies was the improved incentive situation. That is, the workers continued to exert more and more effort in their jobs. In addition, it is possible that the improved incentives led to further increase in skill in the tasks, so that not all of the increased output must be attributed to increased effort. In other words, toward the end of the experiment the workers' capacities may have increased so that they would have produced more even at their original level of effort than they had at the beginning of the experiment. This latter possibility does not appear to have received much attention in the Western Electric Company's interpretation of the results.

If the workers were exerting more effort under the experimental conditions which they liked and which made them feel important, it does not necessarily mean that we should ignore the cost of work altogether. The workers' output improved, but, as we have already pointed out, this may either raise or lower efficiency. The only basis which the investigators had for implying that efficiency was greater in the test-room studies, was qualitative reports. The workers seemed to enjoy their work more. The physical examinations did not reveal any physiological harm resulting from the work, although we should not expect them to show anything unless the work were producing a really drastic physiological upset. To be sure, the workers were told to "work as they felt," and therefore they supposedly felt more like working, as the experiment progressed. It could therefore be argued that if the cost of the work had been great they

[6] Roethlisberger and Dickson, *op. cit.*, p. 127.

would not have reacted in this manner. The detailed reports of the experiments, however, throw some doubt upon the degree to which these instructions were adopted. Two of the original workers were dropped from the group because they insisted on "working as they felt," and so held down the group's monetary bonus.

Thus all the information upon effort and cost of work in the Western Electric studies is vague, qualitative, and inconclusive. This conclusion is not to be taken as a specific criticism of the researchers who were responsible for these studies. It is symptomatic of the state of the field that this study has attracted so much favorable comment, not only among industrial administrators but among psychologists and physiologists as well. The implication is, therefore, that we are not yet ready to carry out anything more than crude and qualitative studies of the "nervous" costs of work. The danger in the Hawthorne studies lies only in the tendency of some readers to fail to recognize the existence of the more fundamental problem of the effect of motivational and attitudinal factors upon the efficiency of the worker.

A more fundamental understanding of "nervous fatigue" will probably come from systematic studies of autonomic function. The crude measures of output, turnover, absenteeism, and worker attitudes can be considered only as rough and ready devices which may aid in practical personnel work. To characterize "nervous fatigue" and to state the directions our search for understanding should take is, however, extremely difficult. The symptoms of headache, indigestion, lassitude, insomnia, tremor, and irritability, all suggest that the autonomic changes which characteristically accompany emotion leave residues in the same functional systems after the work and the predicament are over. It is possible that the autonomic systems themselves become fatigued with prolonged emotional strain, although we do not know the mechanism. The problem is complicated by the fact that we are dealing with systems which are continually active, and which do not require "rest" in the sense that the skeletal musculature requires it. The changes are probably more subtle than the changes in muscular fatigue, although the latter are complex enough, as we have seen.

Emotional Factors and the "Work-Itself"

In the preceding section we have criticized the tendency to overlook the problem of determining the cost to the individual of emotional strain and dissatisfaction. While it has been repeatedly stressed

in current personnel literature that these emotional factors are important, nevertheless these problems have not been integrated with the concept of efficiency. It is true that we are not now in position to work out the interrelations between emotional factors and the cost of work, but we do feel that there is a great need to discover these relationships. The tendency has been, however, to avoid even a statement of the problem.

We now examine another mistaken tendency which is closely related to the one just discussed. This is the assumption that because emotional and attitudinal factors are so much more important than the cost of the "work-itself," the latter can be considered a negligible factor in most industrial tasks. More concretely, it is admitted that the muscular strain and abnormal physical conditions in such an occupation as that of the deep-sea diver, sand hog, or stoker cost the worker a great deal, even if he works under good motivation, does not bicker with the foreman or fight with his fellow workers, feels that his work is important and that he is well paid. When light manual tasks, clerical occupations, supervisory or executive positions are under consideration, however, it is implied that there is no fatigue or other element of cost to be attributed to the performance of the task unless the worker does not like the work, has a feud with his foreman, or becomes involved in some other social and emotional problem.

When we speak of the cost of the "work-itself" we refer to those aspects of cost which correlate with the method of doing the task, or with the surrounding physical conditions which are a natural part of the working situation. While we cannot eliminate emotional factors altogether, nevertheless they can be held reasonably constant in the laboratory and even in the factory, while we investigate the effects of methods of work, temperature, lighting, and the like. What we find objectionable in some current personnel literature is the implication that studies of the effects of work methods or the like are of very little value in comparison to studies of the attitudes of workers toward management and supervision.

The question to be considered boils down to this: What is the value of knowledge of the effect of the work itself in studying the efficiency of the worker under normal circumstances? We believe that this knowledge should have greater importance than is attributed to it in current discussions, particularly those which are based upon the Western Electric Company findings.

Perhaps our point can be made most clearly by an example of a conference which the writer once attended. The research under dis-

cussion had to do with the physical design of an article of household equipment—the optimal height for the work. The researcher had conducted experiments upon the metabolism, pulse rate, and amount of shifting of body weight of a group of subjects as they worked at each of the experimental heights of the article of equipment. Among those discussing the results were several who had been impressed by the importance of emotion in determining efficiency; these continually brought up the fact that the housewife's quarrels with her husband, or her dislike of household tasks, might have more effect upon her efficiency than the height of the equipment. While such a statement is doubtless true, it is completely irrelevant to the question at hand, except as a source of variation which should be controlled by proper experimental precautions. The number of quarrels between husband and wife is not going to vary with the height of the equipment at which the housewife works, and we need not consider it in evaluating the design of equipment.

Of course the situation is not always quite so simple as that cited above. In introducing rest periods, for example, we must consider their effectiveness from the point of view of attitudes as well as work methods. Any rest periods at all may be beneficial if they reduce the worker's resentment toward management or his job. There are frequent cases on record in which this appears to be part of the explanation of the benefits derived from rest periods. Nevertheless the problem of what kind of program of rest and work to adopt must take into account the fatigue developed by the work itself. Rest periods and their optimal arrangement would be an important problem even under ideal conditions of worker motivation. When they serve a double purpose of providing increased incentive as well as relieving fatigue, they should still be designed to produce the utmost relief from fatigue.

In some instances there may even be a conflict between the effects of a factor upon the cost of the work itself and its effects upon the attitudes of the worker. A new method of doing a job may be resented by the workers even if it is known to make the task easier. Under these conditions the net outcome would be difficult to predict, but such a situation should not arise. If the method can be shown to be beneficial to the worker, it is the task of the management to "sell" the method in such manner that the resentment does not arise.

In summary, it does not detract from the importance of attitudinal, motivational, or emotional factors in work if we insist that the cost of the "work-itself" must also be studied and that we should search for methods of reducing that cost. Such studies should con-

tribute to improved attitudes of the worker, as well as to the reduction of the direct cost of work. The tendency to devote exclusive attention to the attitudinal and motivational side is likely to give the reader of current publications the notion that work costs the worker nothing of any importance. It is implied that if we can only make him feel well disposed toward management, and make him like his work, the worker can turn out vast amounts of goods for indefinite periods without any harm to himself. Such an implication is certainly to be avoided. While the attitudinal side is important it is by no means the whole story.

Boredom and Dissatisfaction in Repetitive Work

We have already referred, in the preceding chapter, to the factors which affect a worker's interest in his work. As we indicated there, the principal aspect of this problem which has been studied is boredom and dissatisfaction in simple routine or repetitious tasks. The reason for concentrating on this kind of work is apparently that the more extreme instances of dissatisfaction are to be found in these occupations. In fact, some have assumed that repetitious work is naturally boring, and have decried the tendencies of modern industry to increase the number of jobs of this kind and to reduce the amount of individual initiative which the worker can exercise.

Repetition is, of course, one of the factors which make for monotony. We must not, however, assume that it is the only one, or even that it is the most important factor. As we have suggested before, it is often difficult to separate the effects of the nature of the work from such other aspects as the prestige value, pay, or hours of work. As it happens, these other sources of dissatisfaction often appear along with repetitiousness and so enhance its effect upon the worker.

Another reason why we cannot consider that repetition is the primary cause of monotony is the wide individual variation in attitude of different workers toward the same job. Almost any occupation can be boring for some persons and interesting for others. Of course the *average* level of interest varies from job to job, indicating the effect of factors in the task itself, but wide individual differences are superposed upon these job factors in determining the degree of satisfaction of a given worker.

The general problem is therefore twofold—to discover the job (or situational) factors which tend to reduce the *general* level of interest of the workers in that occupation, and to determine the factors which

make an *individual* especially susceptible to boredom in a given type of work. In our general discussion of motivation, we have sketched briefly the factors which have been analyzed in studying the first of these problems—the job or situational factors. We now undertake a more complete analysis of the topic of boredom, with the principal interest focused upon the individual factors which are very important, and which have received more attention from researchers. First it will be necessary to analyze the general problem a little more fully.

The Nature of Boredom and Its Effects.—Our problem has its origins in common sense, where a distinction is made between fatigue and boredom (monotony). It is possible, of course, that the phenomena are not distinct, and some writers in this field have come to the decision that we cannot distinguish between boredom and fatigue. The worker, however, is able to maintain the distinction without difficulty, and the factors which lead him to report tiredness do not appear to be the same as the factors which result in boredom. It therefore seems better to maintain the distinction until we have definite evidence that separation of fatigue and boredom has no useful purpose. Those who refuse to recognize the distinction are those who insist upon an "operational" definition of fatigue in terms of reduced output. Since boredom also leads to reduced output, it cannot be clearly distinguished from fatigue under this definition. Our distinction is based upon the distinction between motivational and governing factors in performance. Fatigue reduces the capacity for performance, while boredom reduces the effort level—it is a reduced level of motivation of the worker, involving a distaste for the work, a desire to cease work.

There have been a number of attempts to find the correlates of boredom in the behavior and productive activity of the worker and some studies of internal physiological changes which accompany the report of boredom. If these correlates could be established, they would be important not only in our general understanding of the nature of the phenomenon but might also provide us with indices of boredom which are more readily measurable and more reliable than the statement of the worker.

The relation of boredom to the shape of the daily output curve has been the subject of several investigations of the British Industrial Health Research Board. At first they concluded that the output curve for a bored worker was U-shaped for each work-spell, in contrast to the "typical" output curve for fatiguing and interesting work where the high point of production tends to be at the first quarter

or halfway point of the spell.[7] In other words, the "monotony curve" was pictured as being the inverse of the "typical" output curve. The same investigators later decided that there were at least four kinds of output curve representing different combinations of fatigue and monotony.[8] In the extreme case, output rises progressively throughout the working period. (See Figure 21, d.) Here it would be concluded that fatigue is at a minimum, and, if practice is ruled out, the improvement would be due to the improved feelings of the worker as quitting-time approaches. The U-shaped curve would represent a more mixed, but probably a more usual, kind of curve for work that is boring and somewhat fatiguing. (See Figure 21, b, c.)

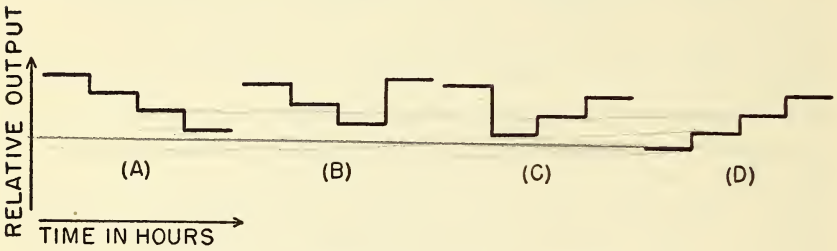


Figure 21. Hypothetical Work Curves

(A) Produced when worker is interested, but suffers from "progressive fatigue."
(B) and (C) Mixed curves. (Both fatigue and boredom.)
(D) "Occurs when the work is distasteful and the operative is severely bored."

(From Wyatt, Langdon, and Stock, *op. cit.*, 1938, p. 29. Reproduced by permission of the Controller of His Britannic Majesty's Stationery Office.)

In order to test this classification of work curves, Wyatt and Fraser classified sixty-eight workers according to the type of work curve which they typically showed. The degree of boredom with the task was assessed for each worker from a questionnaire, and each was rated on an arbitrary scale of degree of boredom. The average boredom rating is shown in Table 17 for each curve type. The significance of the differences in rating was not computed, but the authors state that the classification of the worker by curve type was correct in fifty-one of the sixty-eight cases. This is substantial agreement, when we consider that the classification was made into four types, and that a chance result would show about 25 per cent correct instead of the 75 per cent actually obtained. Such a result would be statistically significant.

[7] S. Wyatt and J. N. Langdon, assisted by F. G. L. Stock, The effects of monotony in work, Industrial Health Research Board (Great Britain), 1929, Report No. 56.

[8] By the same authors, Fatigue and boredom in repetitive work, Industrial Health Research Board (Great Britain), 1938, Report No. 77.

TABLE 17 *

Type of Curve	Boredom†
a	0.9
b	1.7
c	2.8
d	3.6

* From Wyatt *et al., op. cit.*, 1938, p. 30. Reproduced by permission of the Controller of His Britannic Majesty's Stationery Office.

† Calling no boredom 0-1, slight boredom 1-2, moderate boredom 2-3.5, and severe boredom 3.5-5.0, the curve assessments agree with the boredom assessments in 4 cases out of 6, 15 out of 20, 22 out of 26, and 10 out of 16 respectively.

The appearance of a rising, or U-shaped, output curve can therefore be taken as a clue to the existence of boredom of the worker. Other factors may, however, produce similar patterns in the absence of extreme boredom. Link observed steadily rising output throughout each work-spell of shell inspectors, but he concluded that it was due to the desire of the worker to reach a certain goal by the end of the day, to anticipated bonuses for reaching certain levels of output, and, possibly, to the effects of practice and rhythm.[9]

In an unpublished study, Cain kept complete records of output and other variables for each of sixteen workers over a period of a week.[10] These workers were all engaged in sewing operations, and there was considerable variation in their reports of boredom with the work. The individual's work curve varied from day to day, and no correlation between curve shape and monotony could be established. (See Figures 22 and 23.) In fact, few of the curves showed any of the "typical" patterns for either fatigue or monotony. The output during the last hour of the day might show an increase or a decrease as compared with earlier periods. One factor which was important here was the personal goal or quota which the worker had set for herself. In this particular plant, a bonus system was in effect, but the "task" was so high that few workers attempted to reach it, and therefore the bonus was seldom an incentive. Instead, the workers had determined on other bases what they considered to be a normal day's work. If they were behind near the end of the day, they spurted in order to catch up. If they were ahead, they might even cease work altogether, clean the machine, or just potter until quitting

[9] H. C. Link, A practical study in industrial fatigue, *Journal of Industrial Hygiene,* 1919, *1,* 233-237.

[10] P. A. Cain, Individual differences in susceptibility to monotony (unpublished doctoral dissertation), Cornell University, 1942, pp. 58-86.

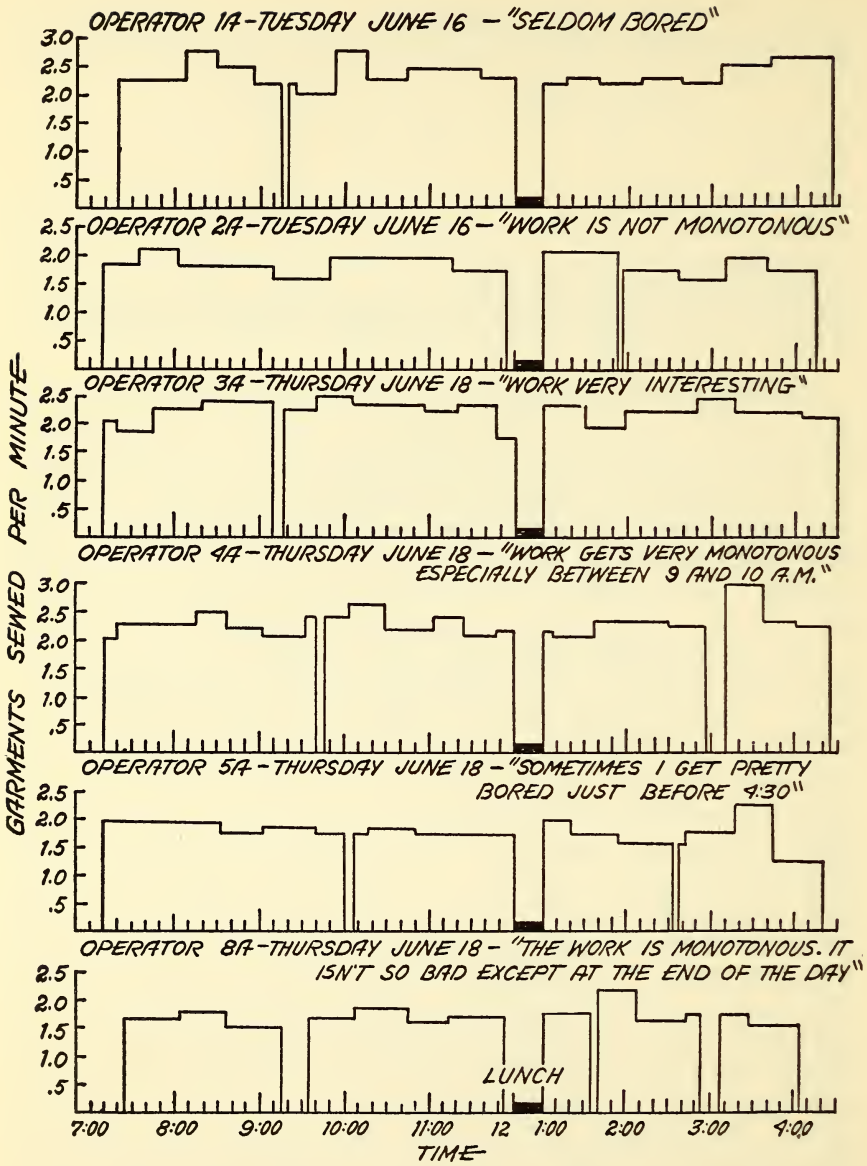


Figure 22. Production Curves for a Repetitious Task. Individual Curves Compared with Workers' Reports on Boredom

(From Cain, *op. cit.*, p. 68.)

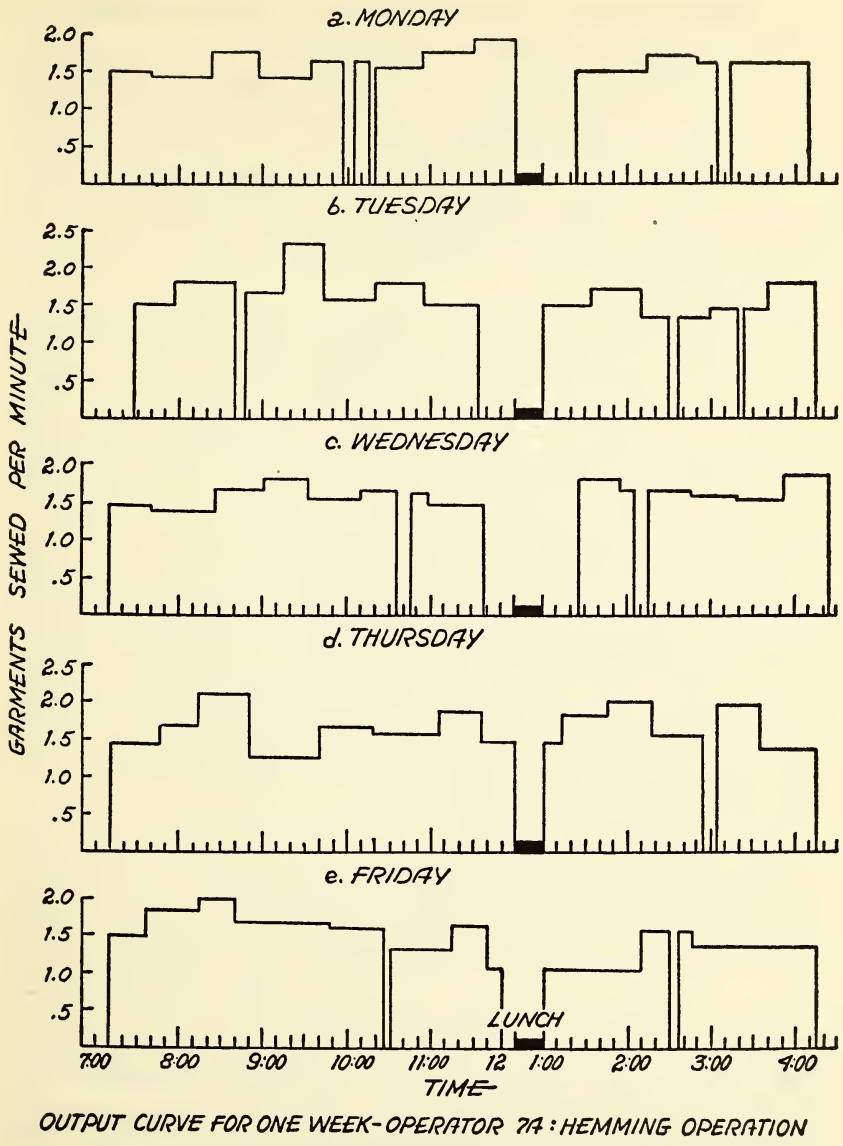


Figure 23. Production Curves for One Week of Repetitious Work—One Worker Who Considers the Work Very Boring

(From Cain, *op. cit.*, p. 67.)

time. Factors like this, rather than degree of boredom, seemed to be the crucial factors shaping the output pattern.

Other behavioral indicators have also been proposed, among which are variability of performance, frequency of talking, and frequency of voluntary rest pauses. Under certain conditions these may be correlated with the degree of boredom of an individual worker, just as the shape of the output curve may be. Cain's study of sewing-machine operators, however, produced negative results for these behavioral indicators as well.

It appears, therefore, that the behavior of the worker can be related to degree of boredom only under certain conditions. These conditions under which the behavioral indicators have some validity cannot, however, be specified. It is also possible that these same indicators are affected by many other factors in addition to boredom. They certainly do not furnish us with trustworthy "objective" indicators of boredom. The only method of gauging the tendency to boredom in a given worker is to question him as carefully as possible.

Here the problem is essentially the same as that in any study of motives or attitudes by means of interviews or questionnaires. A series of specific questions about various aspects of the work is likely to give a more valid indication of the worker's boredom than one or two general questions. Wyatt and his co-workers devised a standardized series of questions which were then scored on an arbitrary scale of boredom.[11] This score was used as their criterion in validating other indices of boredom.

The organic changes which are related to boredom are even more unclear than the behavioral changes. The most extensive study of this problem was that of Barmack, who measured oxygen consumption, blood pressure, and pulse rate during the performance of various "mental" and motor tasks.[12] He searched for possible relationships between the reports of the subject at a given time, his rate of performance, and these organic measures. The relationships, as we might expect, were extremely variable. Barmack's conclusions are based upon a qualitative examination of the data rather than upon a statistical analysis of the significance of the interrelationships. He believes there is a trend toward reduced oxygen consumption and vascular activity during boredom, and an increase in those functions during interested participation in a task.

---

[11] Wyatt, Fraser, and Stock, *op. cit.*, 1938, 4.

[12] J. E. Barmack, Boredom and other factors in the physiology of mental effort; an exploratory study, *Archives of Psychology*, 1937, No. 218.

Individual Factors in Susceptibility to Monotony.—Several researches have been devoted to the search for the determinants of individual variability in the effects of repetitious work. In most of these studies the investigator has been interested in testing some theory concerning the nature of the monotony-susceptible worker. Münsterberg, for example, attempted to show that those who have difficulty in judging the number of similar items in a list of words disliked uniformity in their daily activities.[13] This approach was based upon a theory that some persons are much more sensitive to variation than others, and that this was an important factor in adapting to uniform, repetitious work. Others have considered the ability to automatize a task (to make the performance a "habit" and to be able to think of other things while working), intelligence, introversion-extroversion, the tendency to "perseverate" (difficulty in shifting from one activity to another involving reversed motions) and so on.

Most of the results of these studies have provided little support for the theories under test, and, more important for our purposes, have not provided us with predictive devices which would enable us to sort out those workers who would be more adaptable for repetitious work. One factor in the above list must be considered somewhat more in detail, however, because it has been given a good deal of stress by personnel administrators. This is the factor of intelligence.

There are two kinds of evidence that low intelligence is associated with ability to adjust to repetitious work. From some books on personnel administration, the reader may get the impression that this evidence is conclusive and that intelligence is a primary factor in deciding the suitability of a worker for the simpler manual occupations. In fact, however, this evidence is by no means so clear cut as these textbooks would have us believe. We must therefore consider the results more fully.

One line of evidence is based upon turnover data. In clerical occupations, for example, Bills showed that the turnover rate was much higher for persons of high intelligence in the simpler levels of clerical work than it was for those of low intelligence. The converse was also true. Turnover, however, does not measure boredom. It may well be that persons of high intelligence leave simpler jobs because they are able to find other work which pays higher wages. The value of intelligence as a predictor of boredom is therefore left in doubt.

[13] H. Münsterberg, *Psychology and Industrial Efficiency,* New York, Houghton Mifflin Co., 1913, pp. 195-198.

More direct evidence was obtained by Wyatt, Fraser, and Stock, who related intelligence scores directly to degree of boredom with the work as reported in a questionnaire.[14] Intelligence scores of the "most bored" workers and the "least bored" individuals in each of several groups are shown in Table 18. The small

TABLE 18*

AVERAGE SCORES IN A TEST OF INTELLIGENCE OBTAINED BY THOSE INDIVIDUALS WHO RECEIVED THE HIGHEST AND LOWEST ASSESSMENTS FOR BOREDOM

		Most Bored		Least Bored		Difference
Factory	Group	No. in Group	Intelligence Scores	No. in Group	Intelligence Scores	Between Groups
	A.1	10	20.0	10	15.5	Significant
A	A.2	7	20.4	7	17.9	Not significant
	A.3	7	20.3	7	15.7	Not significant
	B.1	5	19.6	5	12.2	Significant
	B.2	4	26.0	4	19.7	Not significant
B	B.3	5	28.0	5	21.0	Significant
	B.4	4	24.6	4	15.8	Significant
	C.1	6	15.7	6	14.5	Not significant
C	C.2	6	15.8	6	15.5	Not significant
	C.3	4	20.0	4	14.7	Not significant

* *Ibid.* Reproduced by permission of the Controller of His Britannic Majesty's Stationery Office.

size of each group makes it impossible to interpret the results, since the lack of significance may be due only to the small number of individuals in the sample. The authors of the study conclude: "Although the groups were small, and the differences seldom statistically significant, yet in each of the ten groups the average score in intelligence of those who were most bored by the industrial process was higher than the corresponding score obtained by the least bored members of the group. For this reason it is permissible to conclude that the amount of boredom experienced by operatives employed on repetition work increases with their degree of intelligence."[15] Apparently there was so much difference among the different factories and subdivisions of the subjects that the authors did not believe a combination of all of the results would be any more significant.

[14] Wyatt, et al., *op. cit.*, 1938, 19.
[15] *Ibid.*, 19-20.

It should be pointed out that prediction of adjustment would require more than significant differences. It would be necessary to show that there is little overlapping between the groups. For example, there might be a significant difference between the average height of college students today as compared with twenty years ago. Yet we should not expect to get very accurate predictions of the height of an individual merely through knowing that he is a member of the present generation of students.

In addition to the fact that Wyatt, Fraser, and Stock have not demonstrated the predictive value of intelligence by these results, there are other results which fail to agree even with the mild association which they found. Cain, in another part of the study of sewing-machine operators and other employees of a clothing plant, gave an intelligence test to twenty-one workers who had been classified according to their monotony susceptibility.[16] This is not enough, of course, to establish significant relationships, but trends should show themselves, since the individuals represented the extremes of susceptibility to boredom, no members of the middle group being included. Here the difference, although slight and not significant, was in a direction *opposite* to that found by Wyatt, Fraser, and Stock.

Cain's investigation did provide, however, some clues which may lead to more adequate prediction of susceptibility to boredom. She administered a questionnaire containing items dealing with personal habits, attitudes toward work, home environment, and the like, to about seventy workers engaged in repetitious semiskilled work. The criterion of susceptibility to boredom was included in the questionnaire and consisted of questions similar to those used by Wyatt, Fraser, and Stock. These criterion questions were mixed in with the other items concerning the other personal characteristics of the worker. The answers to each item were analyzed separately for their relationship to the criterion. Several groups of questions were not only related but also showed internal consistency among themselves.

From this analysis Cain concluded that the monotony-susceptible worker tends to be less contented with her life in general than the worker who is non-susceptible. The monotony-susceptible also tended to be more poorly adjusted in her relationships to family and the home. The non-susceptible liked better the more routine kinds of housework, and preferred the quieter forms of recreation. She also tended to prefer regular habits of daily activity outside of work. In other words, there is evidence for generality of the trait of monotony-

[16] P. A. Cain, *op. cit.*

susceptibility, something which has been frequently assumed but seldom demonstrated.

A system of weighted-scoring of the questionnaire replies, based upon the statistical analysis just mentioned, was worked out. It was clearly demonstrated that the criterion groups—those used in working out the scoring system—could be sharply separated in terms of their total scores, with a critical ratio of over 10.[17] This is not yet, however, conclusive proof of the predictive value of the questionnaire. It is still necessary that the results be checked with a new group whose answers were not used in establishing the scoring procedure,—a group which answers the questions *before* employment. It must then be shown that their scores correlate with their *later* adjustment to repetitious work.

The problem of the generality of the trait of susceptibility, touched upon above in relation to Cain's study, needs further investigation if we are even to develop adequate predictions of adjustment to repetitious work. To speak of a trait as *general* means that a similar kind of reaction is observed in the same individual in a variety of situations or conditions. Thus, if monotony-susceptibility were proved to be general, it would mean that a person who dislikes one type of repetitious work would also dislike other jobs of a repetitious character, even though the jobs were otherwise quite different. As we have mentioned, Cain's results indicated that monotony-susceptibility on the job is related to dislike for regularity of habits outside of working hours. This would obviously indicate generality of the trait, but there is other evidence which must force us to proceed with some caution in our conclusions.

Wyatt, Fraser, and Stock arranged to have the same group of workers employed in five different tasks in the packing department of a candy plant.[18] Each worker spent one month in each of the five jobs, and her attitude of interest or boredom was recorded from time to time. The results of these ratings are shown in Table 19. The numbers in the body of the table represent an assessment of the "degree of boredom" determined from questioning the workers, with 0 representing "complete absence of boredom" and 5 representing "seldom free from boredom." It will be seen that each worker was more bored by some tasks than by others. Wyatt, Fraser, and Stock conclude: "The limited results in Table 19 provide some evidence that boredom may be specific, but a more comprehensive

---

[17] A critical ratio of 3 or more is considered highly significant.
[18] *Op cit.*, 1938, pp. 12-16.

investigation is necessary before a more definite answer can be obtained." [19]

TABLE 19*

SHOWING THE INCIDENCE OF BOREDOM SYMPTOMS IN THE CASE OF THE SAME GROUP OF WORKERS EMPLOYED FOR MONTHLY PERIODS IN DIFFERENT PROCESSES

Worker	Packing (14 lb.)	Packing (4 lb.)	Staying Boxes	Making Crackers	Bundling Chocolates
A	0.4	1.5	1.9	4.0	4.2
B	0.7	1.1	2.0	4.8	4.6
C	3.5	4.0	1.1	1.8	4.3
D	1.4	2.5	1.8	3.7	4.8
E	1.0	0.3	2.1	3.4	3.2
F	1.1	1.9	4.7	1.0	4.4
G	0.9	0.5	3.2	0.3	4.1
H	3.2	2.6	3.0	1.2	4.5
I	0.8	1.5	1.6	2.0	3.8
J	3.6	1.0	1.5	1.9	4.0
Averages	1.7	1.7	2.3	2.4	4.2

* *Ibid.*, p. 15. Reproduced by permission of the Controller of His Britannic Majesty's Stationery Office.

The difficulty with such an investigation is that it would be necessary to leave each worker on each job for a considerable period of time, and without the worker's knowing that a transfer is to be made later. It may be that the one-month period in the above investigation is insufficient to determine the specificity of boredom. Cain found that in the sewing occupations which she investigated the workers who were new at the jobs were relatively free from boredom. After finishing the learning period, however, the workers tend to show high incidence of boredom. Later, some workers readjust and become more interested in the work again. Susceptibility to boredom should not be measured in an investigation like Wyatt's until this period of shifting adjustment is over. The method of selecting workers who are not susceptible to monotony will be most satisfactory if it will predict which workers will make a long-time adjustment to the kind of work.

From such evidence as is now available, it would appear that there are both general and specific factors involved in susceptibility. Workers will differ from one another in their ability to adjust to

[19] *Op cit.*, 15-16.

any kind of repetitive work, regardless of its specific character. At the same time, different jobs, all within the class of repetitious work, will have varying appeals for the same individual. For the most part, predictive methods will be aimed at predicting the general reaction to repetitious work. It does not appear likely that it would be profitable to attempt to predict differentially for the specific kinds of work. That is, it may be feasible to develop a method of predicting that a given individual would probably make a satisfactory adjustment to the *class* of repetitious jobs. It would not be feasible, ordinarily, to decide whether he would make a better adjustment in running a punch press than he would in simple inspection, unless there are aptitude differences which are also easily measurable. Tendencies to boredom which are related to the specific character of the job will probably always be detected by trial and error, except possibly in cases where it is especially important that turnover on a particular job be reduced to a minimum.

Wyatt, Fraser, and Stock make the point that a very slight difference in the type of work may make a distinct change in the attitude of the worker. If these slight variations from job to job turn out upon more thorough investigation to be much more important than the individual's susceptibility to repetition in general, the possibility of adequate selective devices becomes remote. That is why a thorough investigation of this problem is so important.

Chapter 10

TIME STANDARDS AND RATE SETTING

Even though pay may have been overrated as an incentive to high productivity in the past, it remains, nevertheless, one of the key problems of industrial motivation. Relative pay is more important than the absolute amount of money in the pay envelope, not only because of the economic factor of the value of the dollar, but also because the pay needs to be adjusted so that the worker feels that he is treated fairly in comparison with other workers.

The worker's pay depends directly or indirectly upon his rate of production as compared with some standard of reference. If he is paid by the piece, the price paid per piece depends upon what the worker's time is considered to be worth, and the standard, or normal, speed of work expected from a worker on this job. If he is paid a bonus for production in excess of a certain amount, his pay again depends upon a standard of reference—how much he is expected to accomplish with normal effort. Even if the worker is paid by the hour, increases in hourly pay and promotions depend upon whether or not his performance exceeds the normal rate for that job.

The accurate determination of the standard rate or *standard time* for a job is, therefore, an important phase of establishing effective industrial incentives. The standard time is also of importance to management for other reasons. In planning production, determining machinery and labor requirements, or estimating costs, the normal time for each unit of work in each job is obviously essential information.

We need say little at present about the other important phase of pay determination and rate setting—that is, of the problem of determining value of the worker's *time* when he is working at the expected or normal rate for the job. It is evident that this question involves many non-psychological aspects, and that the question is largely a matter for economic study and political argument. There are, to be sure, certain phases of the problem of determining the value of the worker's time which are psychological; these will be taken up later under the heading of techniques of *job evaluation*

in Chapter 11. Even that discussion will not, however, lead us into consideration of such questions as whether management and labor are receiving their fair share of profits. While the opinions of the worker upon such questions will have important effects upon motivation, the answers to such questions are matters of economics, political philosophy, or ethics, rather than of psychology.

Even if these broad questions of the distribution of profits were answered to the satisfaction of all, we should still be faced with the problem of determining what the normal or expected speed of work should be for a given task if pay is to be an effective incentive for performance. If the expected speed is too high, it fosters resentment and the worker may not produce as much as he would with a more accurate standard. If the expected rate of work is too lenient, the worker either fails to do his fair share of work or he receives pay which is abnormally high compared to other workers who are doing similar work and expending the same effort. If the speeds expected by management are equitably determined, there is greater chance that the worker will adopt these standards as his own goals. If they are not just, the workers develop their own standards which may or may not be the most efficient.

Methods of Determining Time Standards

Although some of the methods by which standard time is determined are spoken of as "scientific," we shall see that all the methods now in use are based upon rule o' thumb and judgment, and that many of the techniques do not appear to have any well-defined aim. Nevertheless it should be remembered that these methods are all that are now available. Their shortcomings merely emphasize the need for more adequate research into the problem—research which is likely to be slow and very difficult in its development. Perhaps such research should be sponsored jointly by labor organizations and management in order to insure acceptance of the results as neutral and unbiased.

Before evaluating the specific methods, we must consider what it is that we want to accomplish. As we have already mentioned in an earlier example in this book, the ideal is that workers on different jobs, but with the same skill, training, and risk, will receive *equal pay for equal effort.* General training requirements, the conditions of work in the occupation as a whole, risk, and the like, are taken into account in deciding the value of the worker's time (job evaluation), as was pointed out above. In setting job standards, it is necessary

to hold skill and effort constant in order to find out how much is produced in each kind of work at this constant level. Standard time, in other words, refers to the time required by a worker of some standard skill and using a standard quantity of effort.

It may be well to analyze the implications of the preceding paragraph in somewhat greater detail, even though at the moment it may appear to involve the unnecessary expansion of obvious statements. Even though obvious, some of these points have been overlooked or confused in the various techniques of time study now in use. First of all, we may accept as commonly agreed that a man should be paid in accordance with his skill. If he is able to work faster than the next man with no greater effort, we consider that he is entitled to the extra gain he can receive in piecework or bonus payments. In addition, two men of equivalent skill should, we believe, receive pay in accordance with their relative amounts of effort exerted in the task.

To say that a man is paid in accordance with both skill and effort may appear to be no more than the statement that he is paid for how much he produces. It is another way of saying it, however, and a separation of the factors of skill and effort is necessary if we are to adjust the relative pay allotted to different jobs and work under different conditions. For example, the statement that pay should be proportional to output does not tell us anything about the *relative* piece rate which should be applied to two different jobs such as drilling castings and operating a milling machine. The problem of the relative worth of one operation on the drill as compared with one operation on the milling machine cannot be solved without some analysis of the factors of skill and effort as separate contributors to performance.

Although *skill* is a term which is fundamental in much of time study, it is seldom clearly defined. We can define the term quite simply in relation to other concepts which have been used throughout this book. An individual's skill simply refers to his capacity for performance of a particular kind measured under some standard condition. It represents the effect of aptitude *plus* training and experience. There are, however, certain restrictions which must be included in the field of time study. If a man is taught an entirely new method of performing his work so that he is really not performing the same job any longer, his capacity for output may be, of course, increased. It is not to be expected that the man should be paid for this increase in capacity for performance. If he is the inventor of the new method he may be reimbursed for his invention, but this is a separate problem.

Some time-study engineers have taken the point of view that any change in skill entails a change in method of work, and is therefore not to be taken into account in determining pay rates. The point of view would be, apparently, that the changes in method could be taught to other workers, and therefore would not constitute a factor for which a man should be reimbursed in his regular earnings. Even if skill were entirely a matter of method of work, which remains to be proved, many of the changes would be so subtle that they could not be laid down as a part of the instructions for the job. Consequently such variations in skill must be taken into account.

We have already discussed the meaning of the term *effort* in an earlier section of this book. (See pages 95-98.) The term appears to be used in much the same sense in time study, although its meaning is more frequently implied than explicitly stated.

We may now formulate the problem of time standards a little more explicitly. The standard for the job is to be the time required for a man of a certain standard level of skill working with a certain standard level of effort. Variations in productivity, whether due to changes in effort, to changes in capacity (skill), or to changes in both factors together are referred to this standard rate of production in order to determine *how much* additional pay should be given for the extra output.

The ideal solution of the problem would be based upon an experimental analysis of the separate contributions of these two variables. The range of skill might be determined for a large group of unselected individuals who work under conditions of fixed effort-level. Some point in this distribution could be chosen as a reference point. Then individuals of a fixed level of skill or capacity would be observed while working at variable levels of effort. Unfortunately, such experimental control and sampling is impossible under practical industrial conditions, due to the vast numbers of jobs which must be standardized and to many other complications which arise. A number of short-cut substitutes for the above ideal procedure have therefore been developed by practical time-study engineers. We must now consider how adequate these substitutes are.

1. Standards Set by Tradition.—In many occupations which have been carried on with little change in method for many years, there is common acceptance of a level of performance which is considered a normal day's work. The brick layer is expected to lay x bricks per day in straight work, the painter can cover about x square feet of

clapboard wall, and so on. While a well-motivated workman ought to be able to determine a reasonable speed of work over a number of years at the same kind of work, the worker in the normal working situation is not motivated to find the most reasonable rate of work (the one which he can maintain without excessive fatigue). Instead, he is motivated to find the lowest rates which will be accepted by the employer.

Even if we were willing to accept rates of performance sanctioned by tradition as satisfactory bases for evaluating a man's performance and setting his pay, this procedure would be applicable only to a relatively small number of jobs, such as the standard skilled occupations. Large numbers of factory jobs change so rapidly that there is no time for a body of skilled workers to arrive at a traditional rate of performance. Consequently it is understandable that industry has sought something more objective, and something which would permit a statement of the standards for a job almost as soon as that job begins. The other methods which we shall consider represent examples of the solutions which have been widely accepted as the best available at the present time.

2. Time Study with Rating.—In order to understand this widely used method, it will be necessary to describe briefly the steps by which the time-study engineer arrives at standard time for the job. He defines *standard time* as the time required by a person of average skill, working with average effort with an hourly wage, and under average conditions, to do the job in question. He also takes into account allowances for decrease in rate of work due to fatigue as the day goes on, normal interruptions in work such as waiting for more material, rest periods, and the like.

It should be noted at the outset that the term "average" is not clearly defined, but that it probably does not mean a statistical average. It is, rather, an arbitrary standard or norm by which the time-study engineer works, and certainly it does not coincide with the average level of performance of workers on the job. Barnes, for example, gives a table illustrating individual differences by showing the distribution of rates of performance in a group of workers on the same job.[1] In this group, only four of the one hundred twenty-one workers take longer than the standard time. Of course they were presumably paid on the basis of a piece rate or bonus, while "standard time" refers to performance without this additional financial incen-

[1] Barnes, *op. cit.*, p. 271.

tive. It is difficult, therefore, to determine whether the standard is intended to refer to a true average under the latter conditions or not. At any rate, some caution must be applied to the use of the term *average* in this connection.[2]

The standard time can be determined with the rating or "leveling" method from observations upon a single worker, although the results may be checked by observations upon other workers. The time-study man first observes the job and lists the component parts or elements which make it up, describing them as "place part in jig," "tighten nut," "push hand lever," and so on. After making certain that he has listed all possible parts of the operation, he times a worker with a stop watch, with the knowledge and consent of the worker concerned. The time for each part of the operation is recorded separately.[3]

If a sufficient number of records is made, it is possible to derive the average time for each part of the job and for the job as a whole—for *this* worker, under these specific conditions, with a given amount of effort. It is now necessary to make allowances for the special worker and conditions in order to *level* the times to the *standard* worker and standard effort and conditions.

The procedure of "leveling" developed by the Westinghouse Manufacturing Company is based upon ratings.[4] The time-study man rates the worker's skill, effort, and consistency, and the condi-

---

[2] Shumard states that production should average one-third greater under a suitable piece-rate system than under a straight time payment. (F. W. Shumard, *Primer of Time Study,* New York, McGraw-Hill Book Co., Inc., 1940.) If this is true, average time under piece rate would be only three-fourths of standard time, even if workers of "average" skill are employed. If the workers are selected for skill, there would be a still greater reduction in average time. The problem is complicated by the fact that piece-rate systems will vary in their effect, depending upon the type of work and other factors. (Cf. S. Wyatt, Incentives in repetitive work, Industrial Health Research Board (Great Britain), 1934, Report No. 69.) As the standard time is defined, it is therefore difficult to see how useful it would be for work scheduling or other situations where it is desired to predict truly average production rates.

A more extended analysis is given in a recent book by R. Presgrave (*The Dynamics of Time Study,* (2nd. Ed.), Toronto, University of Toronto Press, 1945) where it is recognized that the standard is to be adjusted in a manner to make the pay incentive maximally effective. Presgrave chooses to adjust the standard so that average performance of a group of workers would be approximately 25 per cent to 30 per cent above standard when paid on a piece-rate plan. Only about 5 per cent of the group would fall below "standard."

[3] The reason for timing each part of the job separately, rather than taking the over-all time, involves other aspects of time study which we do not need to discuss at the present time.

It is understood here that the proper method of doing the job is already established. See Chapter 7 for a discussion of procedures by which methods are developed—motion study or methods engineering.

[4] S. M. Lowry, H. B. Maynard, and G. J. Stegemerton, *Time and Motion Study,* (3rd Ed.), 1940, McGraw-Hill Book Company, Inc., New York, p. 233.

tions under which the study was taken. These ratings are translated into numerical "leveling factors" which are used to correct the observed times. Table 20 shows the descriptive ratings which are employed in this system, and the corresponding numerical factors.[5]

TABLE 20*

LEVELING FACTORS

Skill			Effort		
+ 0.15	A1	Superskill	+ 0.13	A1	Excessive
+ 0.13	A2		+ 0.12	A2	
+ 0.11	B1	Excellent	+ 0.10	B1	Excellent
+ 0.08	B2		+ 0.08	B2	
+ 0.06	C1	Good	+ 0.05	C1	Good
+ 0.03	C2		+ 0.02	C2	
0.00	D	Average	0.00	D	Average
− 0.05	E1	Fair	− 0.04	E1	Fair
− 0.10	E2		− 0.08	E2	
− 0.16	F1	Poor	− 0.12	F1	Poor
− 0.22	F2		− 0.17	F2	

Conditions			Consistency		
+ 0.06	A	Ideal	+ 0.04	A	Perfect
+ 0.04	B	Excellent	+ 0.03	B	Excellent
+ 0.02	C	Good	+ 0.01	C	Good
0.00	D	Average	0.00	D	Average
− 0.03	E	Fair	− 0.02	E	Fair
− 0.07	F	Poor	− 0.04	F	Poor

* Reproduced by permission from S. M. Lowry, H. B. Maynard, and G. J. Stegemerton, *op. cit.*, p. 233. Copyrighted, 1940, by the McGraw-Hill Book Company.

Let us illustrate the application of this rating procedure by a crude example. For simplicity, we shall assume that the factors are to be applied to the time for the job as a whole, rather than to the individual parts of the job.

[5] Other rating methods appear to be simpler than the Westinghouse method, but they are actually based upon the same general ideas and are simpler only because they are less explicit in stating precisely what is involved. See F. W. Shumard, *Primer of Time Study,* New York, McGraw-Hill Book Co., Inc., 1940, and Phil Carroll, Jr., *Time Study for Cost Control,* New York, McGraw-Hill Book Co., Inc., 1938, for examples of other rating procedures.

TABLE 21

To Illustrate the Application of Leveling Factors

Average time for worker under study	10.00 minutes
Ratings of worker	Leveling factors:
Skill: Fair (E1)	−0.05
Effort: Poor (F1)	−0.12
Conditions: Average (D)	0.00
Consistency: Poor (F)	−0.04
Total of leveling factors	−0.21
Time for average worker (21 per cent deducted from 10.00 minutes)	7.90 minutes
Fatigue and personal allowances	10%
Standard time (10 per cent added to 7.90 minutes)	8.69 minutes

The reader will probably raise at once the question of how accurate these ratings may be. Other questions will probably occur to him—such questions as why consistency should be considered, since it is probably already taken into account in rating skill and also affects the actual times recorded in the study. Also, how was the numerical factor determined for each rated level of skill, effort, or conditions?

We have chosen to illustrate the details of only one out of several methods in use. The proponents of other methods criticize the above procedure in terms of certain of its specific aspects, but all methods have many features in common. Although the term "leveling" usually refers to the method of Lowry, Maynard, and Stegemerton, sketched above, we shall find it convenient to refer to all similar procedures as "leveling methods," even though the authors of these other methods would object that their procedures are distinct from the method just described. The reason we feel justified in this grouping is because most of the considerations which we shall discuss in succeeding sections apply with slight variation to all procedures involving rating in finding the standard time.

As an example of another procedure which is believed by its author to be more accurate than the Westinghouse method, we may mention the effort-rating method of Presgrave.[6] Presgrave restricts the correction which he applies to time-study data to a rating of effort alone. His position seems to be that skill cannot be indicated upon a numerical scale insofar as it involves variations in method of

[6] R. Presgrave, *op. cit.*

work. Variations in speed which are dependent upon practice or aptitude will appear as inseparable from the influence of motivation or effort. Consequently Presgrave uses an effort rating which he makes synonymous with the rating of "speed" or "tempo," and which would presumably include any purely quantitative effects of skill. If skill entails variations in method, Presgrave believes that it is outside the province of time study to settle the issues involved.

The result of Presgrave's analysis is a confusion rather than a clarification of terminology. It may be that in practice his method is as reliable or even more reliable than others. That is a question which can be settled only experimentally, and in succeeding sections we propose to consider questions of the accuracy of time study in greater detail. Presgrave's claims for the method are based, however, upon its logical consistency, and he presents little concrete data upon its empirical accuracy. His method is essentially equivalent to the leveling method, since it involves a rating factor for the composite which he calls *effort,* a rating factor which is used to correct or "level" the observed time-study data.

Reliability of Leveling. The answer to the detailed questions about the method seems to be that the technique was evolved through trial and error, and the various factors adjusted by experience. As for the general question of accuracy, direct evidence is scanty and unsatisfactory. The problem must be broken down into two main parts, just as any problem of measurement, estimation, or prediction must be. The first phase of the question is whether the method is reliable—will different time-study men arrive at the same standard time for the same job?

Without giving specific data, Barnes summarized the opinion of many time-study engineers on reliability as follows:

> Time standards set by any one of a group of competent and well-trained analysts, or by the same one on different days, should fall within 5 to 7 per cent of the average of the group. That is, no one should differ more than 10 to 15 per cent from any other. Some motion and time-study analysts claim that they can consistently do this within 2 or 3 per cent, but this is doubtful.[7]

Barnes does not say whether these values which he considers as estimates of the accuracy of time study would be obtained if different workers were studied in the different time studies being compared, or if they apply only to different studies of the same individual. He

[7] Ralph M. Barnes, *op. cit.,* p. 277.

also does not state whether the analysts were trained together and by the same instructors, or whether they learned their techniques more independently. With these questions still before us, and with the lack of concrete data which now exists, we cannot now draw definite conclusions concerning the reliability of the leveling method.

The question of reliability of the leveling method will need to be separated into two parts before a clear answer can be found. There is, first of all, the problem of consistency of the method as applied to the *same worker* by different time-study men, or the same time-study man at different times. This consistency depends, primarily, upon the reliability of the ratings for skill, effort, and the like, although variations in the man's performance from day to day might complicate the problem when the studies are made at different times. The generally low reliability of merit ratings is well known, and Uhrbrock [8] gives an example of the low reliability of ratings of quality of work. These facts should raise serious doubts of the possible reliability of ratings of effort and skill. The time-study ratings are, however, different in character from these other ratings, so that consistency might turn out to be more satisfactory in the case of time study. Direct study is necessary, and the low reliability of other ratings emphasizes the importance of such research.

The second phase of reliability involves the consistency of time standards when arrived at through studies of *different* workers, either by the same or different rate-setters. Consistency in this respect depends not only upon the consistency of the ratings as discussed above, but also upon the correctness of the numerical leveling factors assigned to the various men under study. In order to achieve consistency of time standards under these conditions, the raters must be able to place the men in their correct relative positions upon the scale of each factor, and the numerical factors permanently assigned to each rated level must be correct. In other words, it is not only necessary that the same worker should always be placed at the same point on the scale of, say, skill, but that two workers who are actually 20 per cent apart in their skill should always be placed that distance apart in terms of the leveling factor assigned. It should be evident, therefore, that consistency of standards based upon the same worker in repeated studies would not guarantee consistency when different workers are involved.[9] Uhrbrock has also considered this phase of

---

[8] R. S. Uhrbrock, A psychologist looks at wage incentive methods, American Management Association, *Institute of Management Series,* 1935, No. 15, 17-20.

[9] The use of the term *reliability* in this connection may be questioned by those familiar with its usage in the field of psychological testing. Reliability involves

the problem, pointing out the wide variation in estimates of the range of individual differences in ability which would make it difficult to establish any fixed constants for use as leveling factors.[10]

These two phases of reliability do not appear to be clearly distinguished by many of those who have considered the problem. The Society for the Advancement of Management has a committee studying the whole problem of reliability.[11] Their present plans are concerned primarily with reliability in judging the same worker, although they do mention a possibility that they will also study reliability of times derived from different workers. Other statements of the problem, like that of Barnes quoted above (page 217) and of Louden [12] fail to point out the distinction.

The kind of research necessary to evaluate the reliability of time standards from this second point of view is self-evident, once the question is raised. Evidently we need an extensive collection of data upon the time standards for identical jobs as established by different raters in different plants. With sufficient data of this kind on a variety of different jobs, the range of error of the method could be clearly established.

These are questions which cannot be answered in the light of data which have been presented thus far. Barnes' statement quoted above is apparently an estimate or opinion without concrete statistical evidence to support it. In fact these questions are seldom even considered in writings upon these methods.

Cohen and Strauss have given us one of the few direct empirical studies of this problem of reliability of the leveling method as it is applied to different workers.[13] Twenty-one operators on the same job were subjected to a motion-picture time study. Their task consisted of "the folding of an 18 × 18 inch gauge sheet to a size of approximately 4 × 4 inches. This was accomplished in six foldings, although micromotion analysis revealed that all of the operators did

---

a problem of consistency of result, but in this case it depends upon the correctness or *validity* of certain of the tools employed by the time-study engineer. Whether it should be termed validity or reliability is therefore a matter of choice. As will be seen presently, there are also other phases of the validity of the method, however, and it seems clearer to reserve the term for this later discussion.

[10] *Op. cit.*, pp. 15-16.

[11] J. M. Juran, Progress report of committee on rating of time studies, *Advanced Management,* 1941, *6,* 110-115.

[12] J. K. Louden, Management's search for precision in measuring a fair day's work, *Advanced Management,* 1942, *7,* 29. Louden suggests using multiple judges in order to improve the reliability, and the rejection of standards when the raters fail to agree, but he does not specify that the different studies be taken upon different workers.

[13] L. Cohen and L. Strauss, Time study and the fundamental nature of manual skill, *Journal of Consulting Psychology,* 1946, *10,* 146-153.

not make their folds in precisely the same manner." [14] Each operator was rated for skill and effort by three trained time-study engineers. Cohen and Strauss analyzed the times taken by each operator, and then applied the leveling factors corresponding to the engineers' ratings.

If the theory of the leveling method is correct, the "leveled time" should be the same, no matter which operator was used for determining it. Actually, the leveled times were not even approximately equal, as shown in Table 22. The maximum value is more than twice as great as the minimum value.

TABLE 22*

TIME STUDY OF OPERATORS' PERFORMANCE IN A FOLDING JOB

Operator No.	Leveling Factor	Actual No. of Frames Taken	(L. F. × A. F. T.) Leveled Time
17	123	48	59.04
1	111	81	89.91
10	108	83	89.64
20	105	93	97.65
16	100	99	99.00
7	86	145	124.70

* Cohen and Strauss, *op. cit.*, p. 148.

In analyzing the film records, Cohen and Strauss found that the faster workers had developed a pattern of movements different from that of the poor operators. Thus they were not only faster in performing their movements, but they also had fewer motions to make. This is probably the main reason for the breakdown of the leveling procedure. A fixed rating procedure cannot take account of these subtle changes, since the opportunities for such improvements will vary from job to job. As we have already noted, Presgrave and others take the position that rating is not supposed to allow for differences in *method* of work. The logical outcome of such an attitude would be that time study cannot be used reliably until a micromotion study has been made of the jobs, and until the method of work has been specified in the most minute detail. This is only seldom done in practice, and even if it were done, the method of one worker would never be exactly like the method of another worker.

Cohen and Strauss's data apply to only one job. It would therefore be dangerous to conclude upon the basis of this one result that all time study is unreliable. Their explanation of the results is,

[14] *Ibid.*, 147.

however, the sort of variation in movement pattern which could be expected in most jobs. We may therefore expect further studies to agree with their findings. At any rate, there is little documented evidence *for* the consistency of the leveling method, and here is concrete and detailed evidence against it.

Relative Validity of Time Standards. A method of setting time standards might be consistent for each specific job and still give incorrect results. That is, even if studies of different workers on the same job lead to the same time standards for that job, the standards may be too high or too low with respect to the standards for other tasks. This question of the proper relative standards for different jobs is referred to here as the *relative validity* of the method. Beyond that there is a still broader question of the *absolute* validity of the standards which will be considered as a separate problem.

The essential problem in relative validity is whether standards of judgment remain fixed from one kind of task to another. It is conceivable, for example, that what is judged as "normal" effort on one job may actually involve much more cost to the individual than what the time-study man considers to be "normal" effort in another task. There may be illusions which make a man in one kind of work appear to be working harder than he is, and which produce the opposite effect in other kinds of work. If these illusions affected all time-study men in about the same manner, there would be no indication of their effect in the consistency of results.

If such variations in standard from job to job exist, they probably constitute the basis for many of the complaints and criticisms of time study by the workers. The possibility of such variation does not seem to occur to the proponents of various time-study methods, however, so that the problem has not been discussed.[15]

---

[15] Presgrave, who supports a method based upon rating of effort or "speed" of work, gives examples to show what is meant by "normal" effort. (R. Presgrave, Effort rating, *Advanced Management,* 1939, *4,* 126-133.) These examples include walking at three miles per hour, typing fifty words a minute, and the like. By such examples it may be possible to establish more reliable criteria in various raters by making the "normal" standard more concrete. How it is determined that walking three miles per hour and typing fifty words per minute require the same effort is not clear. He does point out that the professional record in typing for one hour and the five-mile walking record both represent speeds that are 180 per cent faster than the normal speeds given above, and he believes that these agreements support the validity of his standards. Presgrave's standards are complicated by the fact that he makes no separation between skill and effort, unless skill entails a difference of method of work as well as differences in speed of motion.

Even if we accept Presgrave's standards for walking and typing, these are only examples to guide the time-study man. He must attempt to apply these standards to many other tasks, and in this application the variations we have suggested above could readily occur.

As far as the present writer can see, there is no way in which the relative validity of time standards can be checked until it is possible to apply direct measures of effort and cost of work to subjects performing a variety of tasks. If subjects on various tasks who are working at what the time-study analysts consider "normal" effort are all found to be actually exerting the same amount of effort as directly measured, then the relative validity of the standards is established.

While we can come to no conclusions upon this point, and cannot find even suggestive evidence to consider, the problem of relative validity must be emphasized. It must be realized that establishing the reliability of these techniques is still a long way from demonstrating the fairness of these standards. The studies by the Society for the Advancement of Management, referred to above, are therefore only a first step toward evaluating these procedures.

Absolute Validity of Time Standards. The analysis of the problem of time standards has been phrased up to this point in terms of reliability and relative accuracy of the method. There is another broad question, still more difficult to formulate and to answer, which must be mentioned in order to round out the statement of the problem. This additional aspect of the accuracy of time standards might be termed the absolute validity of the method.[16]

Suppose, for example, that it has been agreed (miraculously) by all concerned what the average earnings of workers in each type of work should be. The standard times determined by leveling might bear the proper relation from job to job (relative validity), but it is still possible that *all* of the times are too lenient so that average earnings on all of the jobs are too high, or the workers are not exerting a reasonable amount of effort. On the other hand, all of the standard times could be too short, so that the effort would have to be unreasonably high in order to bring the average earnings to the agreed level.

There is no way to investigate this broader phase of validity by analysis of time-study procedures alone. Conclusions upon this problem depend upon answering such questions as, "What minimum level of skill is to be considered allowable in each kind of work?" or "What level of effort can we reasonably expect of the person of acceptable skill in each kind of work?"

Absolute validity depends upon two general phases of time-study technique. First, it depends upon the absolute reference points of

---

[16] See S. Barkin, Labor views the working day, *Advanced Management*, 1942, *7*, 32-37.

"normal" effort, skill, conditions, and consistency adopted by the time-study engineer. Even if time-study men can agree upon their use of these standards of reference, they are still purely arbitrary. For example, what the time-study engineer classifies as "normal" effort bears no necessary relation to the optimal effort for the given type of work. In Lowry, Maynard, and Stegemerton's descriptions of the levels of effort, average seems to be an artificial mid-point between the extremes of "poor" effort and "killing" effort. Average effort itself is not described, except with reference to other levels. The criteria of poor effort include such descriptions as "obviously kills time," and "works slowly and appears lazy." Killing effort involves a worker who "extends himself to [a] pace impossible to maintain steadily." Between these two extremes, "fair," "average," "good," and "excellent" effort are spaced at varying intervals.

Whether the arbitrary mid-point of this scale can be set up as a goal, so that workers who work at this pace can feel that they are doing a reasonable minimum of work, is impossible to say. What this question involves does not seem to be considered by those who design methods of time study.

To use a concrete illustration, we have referred on an earlier page to Presgrave's adoption of three miles per hour as a "normal" rate of walking.[17] There seems to be no particular reason for adopting this rate rather than two or four as normal, and from the point of view of fair *relative* treatment of various jobs it would not matter which level is arbitrarily chosen as standard. Suppose, however, that a man is paid on a "piece rate" of $.25 per mile, so that the standard earning is $.75 per hour. The man might be capable of walking at an average pace of six miles per hour, but would hesitate to double the standard earnings. His reasons might involve not only the fear of eventual reduction in pay rate if he maintains this more rapid pace, but also the feeling that he has "done enough" when he has increased the pace by one third or one half.

Golden and Ruttenberg, who are officers in the C.I.O., present the point of view that management has failed to foster adequate morale in the workers with the result that the workers have frequently been successful in obtaining very lenient time standards.[18] Ruttenberg cites one example in which he was able to double the production of a department in which management believed there were adequate stand-

[17] Presgrave, *op. cit.*

[18] C. S. Golden and H. J. Ruttenberg, *Dynamics of Industrial Democracy,* New York, Harper & Bros., 1942.

ards. Of course, any organization which is operating efficiently would have a reserve capacity which could be tapped under high incentive. This would not necessarily indicate that the operating speed should be increased. When it is possible to more than *double* the output, however, one may wonder whether the normal pace fostered by time standards is not too lenient. Golden and Ruttenberg are speaking, to be sure, of wartime conditions. Whether they would suggest that the pace of production could be permanently increased without harm is not indicated.

The ideal standard for any job should come as near as possible to the following requirements:

1. The standard should be low enough so that most workers who are adequately selected for the work could reach standard. If the standards are so high that few workers are able to reach them, the bonus pay system would have little effect, and workers would adopt their own standards of work.
2. The *optimal* rate of work from the point of view of worker efficiency should not be very much faster than the standard rate. Since the standard itself serves as a goal and point of reference, workers will require very high incentive to go very far above standard. This requirement evidently implies a study of efficiency of performance as a function of rate.

These requirements are phrased in terms of a standard which is considered as a minimum standard, such as most time-study procedures give us. It might be well to couple investigations of the reliability and validity of these standards with studies of the incentive value of standards. More complete knowledge of how standards are regarded by the worker, and how they affect his performance, would be very valuable. It might be found, for example, that standards would be more effective if stated in terms of an expected average performance of adequately selected and trained workers, working under current wage incentives. This is in contrast with present procedures which give us a standard well below this *expected* average level. It has certain potential advantages which should be explored. A worker who meets the standard will know that he is actually performing at the optimal rate for a worker of average skill. If there is any tendency to stop when standard is reached, this would be the best rate which could be set. Where scheduling and planning of work are based upon standard times, these actually expected times should be even more valuable than the present minimal standards.

Fatigue and Personal Allowances in Relation to Validity of Time Standards. It was noted above that the standard time as established by time-study methods includes additional allowances for "fatigue," interruptions, delays, rest periods. Usually, the time determined by the leveling process is considered the *best* time which can be made by the average worker; that is, the time required when he is fresh and when he is not interrupted. When that time has been estimated, an additional percentage is added to this base. This percentage is supposed to correct the standard time to a value which would make it representative of what the average worker could maintain throughout the working day, day in, day out.

Even if we were to accept the whole procedure up to this point, the absolute and relative validity of time-study methods depend upon the accuracy of this "fatigue and personal allowance." In the more painstaking time-study programs, a table of allowances is worked out from daylong observations of a variety of jobs. The total time spent by the worker in resting or being away from his work is averaged. The amount of decrement in output late in the working day is computed from the continuous series of time studies. From this data a table of allowances for various classes of work is evolved. Figure 24 is an example of such a table. When additional jobs are to be timed, the allowances to be made for fatigue and personal requirements are read from the table. Other delays must, however, be determined by direct observation.

Time lost from fatigue and personal needs is highly variable from person to person. The aim is to determine an average value. Those who require less time for rest or who fatigue less rapidly are then given additional rewards for their steady accomplishment. As Lowry, Maynard, and Stegemerton point out, stamina and physical fitness would be one aspect of the general suitability of the worker to the work. These should be rewarded, along with skill and effort.[19]

There can be little criticism of the general aim of fatigue and personal allowances, and there is no question that these allowances should vary from one kind of work to another. We are given very little statistical data, however, upon the methods by which these allowances are determined. It would appear that the allowances are in practice closely correlated with the amount of muscular effort involved in the job. While decrement, as we have seen (Chapter 3), is probably greatest when the work is primarily muscular, the cost of work to the worker is not necessarily correlated with the decrement. To make

[19] Lowry, Maynard, and Stegemerton, *op. cit.*, p. 154.

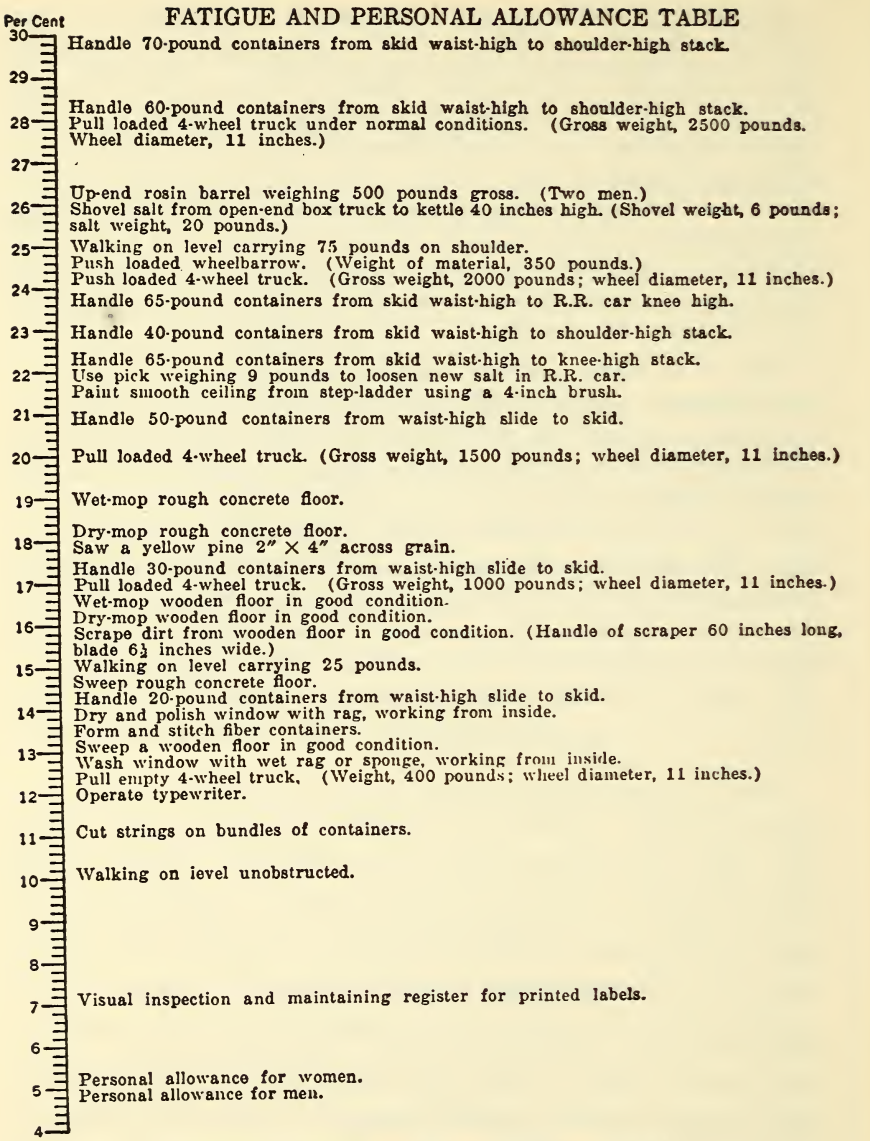


Figure 24. Fatigue and Personal Allowance Table

(Reprinted by permission from *Motion and Time Study* by Professor Ralph M. Barnes. Published by John Wiley & Sons, Inc., 1940. P. 279.)

a large "fatigue" allowance for heavy muscular work, and little or no allowance for light or "mental" work, may result in demanding a higher level of effort for the lighter tasks than that required for the heavier tasks. In other words, eyestrain, concentration, postural

strains, and monotony of the lighter type of work might cost the worker more in the long run than the heavier physical effort which is given a liberal fatigue allowance, even though the lighter work does not show a clear-cut decrement as the day progresses. Once more we see the need for facing the problem of cost of work in its broader aspects.

Proposals to Avoid Rating. Some union organizations have made a proposal which is apparently designed to overcome some of the possible errors of the rating method. They suggest that union contracts specify that an average man should be selected for the time study, presumably eliminating the need for a leveling factor for skill. This proposal is unsatisfactory in several ways, and certainly it does not necessarily work in favor of the worker. In the first place, rating for skill is still involved in the selection of the worker. By requiring that the selection meet the approval of the workers, the rating no longer depends upon the time-study man's judgment, and the workers apparently feel that this would result in a lower standard. This is not necessarily true, however. Suppose, for example, that the workers in a certain job have been carefully selected so that the skill level is extremely high. All the men should receive above-average pay under these conditions. Yet if the skill rating is eliminated, and if an "average" man is selected from the group now working on the job, the men on that job would receive average pay for that kind of work, regardless of their high standards of skill. Whether or not such a situation would often occur, lack of definition of what constitutes an average worker would make consistency of time standards impossible.

Not only does the selection of an average worker for time study fail to eliminate the real problem of skill ratings, but it also fails completely to take into account the equally essential rating for effort which would apparently still remain in effect with the problems of reliability and validity still unsolved.

Another proposal is to return to the use of "average" in a statistical sense. The average would not be an arbitrary standard of the time-study man, but would be the actually observed mean performance in a group of workers on the job. This proposal is essentially a return to tradition or experience as a method of establishing time standards, with the additional feature that the experience which is used as a basis for the fixing of times would be more carefully analyzed and would probably be more accurate.

Such a procedure cannot be regarded as a substitute for a rating method, however, since it fails to accomplish some of the things which a rating method is designed to do. Taking average times from

workers now employed, and under present conditions, merely fixes present levels of skill and effort as standard. Yet there may be, at present, very poor selection of workers or workers who are exceptionally high in skill, or the effort level may be very low as a result of restrictive practices, and as a result of the knowledge that rates are to be fixed by current performance. Also, it would no longer be possible to set rates upon jobs employing only a few workers, nor would it be possible to determine rates for new jobs as soon as they begin. Since a rating method is aimed directly at overcoming all of these difficulties, the proposal to use actual average performance as a standard is merely to abandon the aims of the leveling procedure altogether.[20] This would not be justified unless it had been demonstrated that the rating method was of such low reliability that it really was not accomplishing what it set out to do. Even then, some method for arriving at these goals should be sought.[21]

The statistical average might be employed in research studies to improve the rating procedures, even though they could not be used directly in time study. For example, an experimental study of a job might be carried out under artificial conditions in which the subjects are provided with very intense motivation so that they work nearly at their maximum possible rate. In this way, effort would be held reasonably constant at a high level. Then the distribution of performance in various workers would represent varying degrees of skill, and it would be possible to obtain an average time for the given high effort level. It would also be possible to determine the range of variation of skill by studying the dispersion of the distribution. This would, of course, be valuable information for verifying the leveling factors for skill, although a great variety of jobs would have to be studied in this manner, with a large number of workers in

---

[20] I. Lorge (A statistician looks at time study, motion study, and wage incentive methods, American Management Association, *Institute of Management Series,* No. 18, 18-26) appears to advocate the use of statistical averages in determining standards, but at the same time to retain the leveling procedure. This seems to be a contradiction, or at least an ambiguity in his proposal, as the above discussion would indicate.

[21] W. Bloch (Die Anwendungen der Mathematik in Zeitstudienwesen, 7th International Management Congress, Washington, 1938, *Production Papers,* 144-151) proposes the use of the average time plus a correction factor based upon the standard error of the mean. He would apparently apply this procedure either to a single worker or to groups of workers, however, and his discussion does not appear to take into account the need for leveling for skill. The correction factor he proposes is stated to be a correction for "intensity" of work or effort. The dispersion of times increases as the effort decreases. Since the standard error of a mean depends upon both the dispersion and the size of the sample, he chooses to take this as the corrective factor. Even if this assumption could be fully accepted, there is still no reference to variations in skill.

each task, in order to investigate the variations in range of skill from job to job.

The study of variations in effort could be based upon a prior study of skill like that described in the preceding paragraph. Once workers had been classified into groups of equivalent skill by the procedure described, the workers could then be studied while working at various submaximal levels. Such study might take a variety of directions. More exact criteria of the various levels of effort might come from descriptive studies. For example, the variability of times, the frequency of errors or false movements, or other objective factors might show a relation to the effort level. Effort level itself would be established objectively as it has never been done in time-study work, because the capacity of the individual worker would already be established with some accuracy. In addition, if long-term physiological investigations could accompany such experiments, some hints as to the level of effort which is optimal might be produced. If optimal effort is once established, the problem of absolute validity of time studies can be solved.

All these suggestions are related, however, to the use of statistical averages in fundamental research for improving time study. For the reasons outlined before, averages taken in the ordinary conditions of work would not solve the problem of time standards, since their use would merely avoid the issues by ignoring them.

It should be said that there is some justification for the use of rating methods, in spite of all of the open questions. The developers of the various methods have replied to criticisms by stating that rating in some form is the only reasonable method now available, and that time standards must be determined somehow. There can be little quarrel with this attitude. The danger lies with those users of ratings who regard them as final solutions, and who forget that they are no more than stopgaps. Even though rating must be used for lack of anything better, we should keep before us its provisional nature, in the hope that industry will provide opportunity and support for research designed to test and improve the technique.

3. Synthetic Time Standards.—We have given such a lengthy consideration to the rating methods of time-study because of their wide use, and also because it brings out the fundamental problems of all methods of establishing time standards. As a result we can now deal with the "synthetic" method in a small space.

The synthetic method consists in using data from previous time studies to establish the time standard for a new job without using the

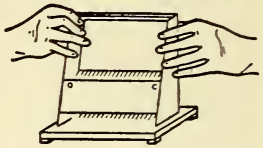
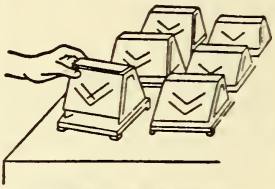
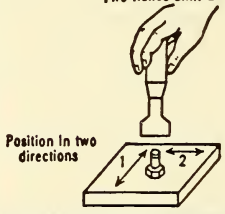
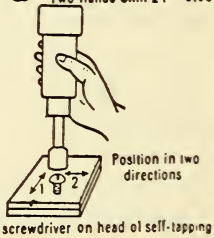
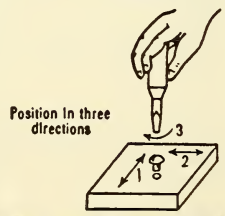
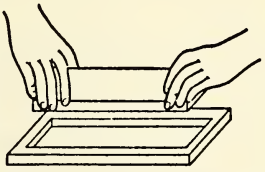
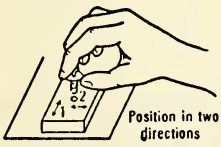
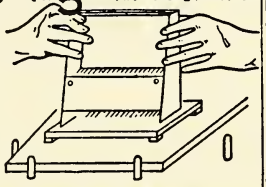
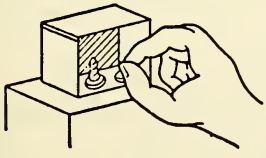
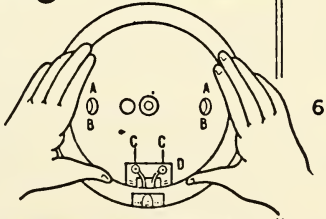
CONDITION OF PLACE Amount of positioning required	SIZE OF OBJECT AND TYPE OF GRASP (Medium) 3 F Three fingers and thumb	 (Large) H Extended hand
CONDITION A Positioning is normally little more than releasing the object on the work place.	Time Class No. 1 One Hand 8″ = 0.006 One Hand 12″ = 0.007 Two Hands Sim. 12″ = 0.011 Place small gear blank in other hand.	Time Class No. 1 One Hand 8″ = 0.006 One Hand 12″ = 0.007 Two Hands Sim. 12″ = 0.011  Place toaster body in other hand for subsequent adjustment.
CONDITION B Positioning of parts on or into definite locations with ample tolerances, simple open nests or fixtures, or assemblies with one point of location.	Time Class No. 1 One Hand 8″ = 0.006 One Hand 12″ = 0.007 Two Hands Sim. 12″ = 0.011 Dispose of screwdriver into funnel-type holder 8 inches to right of work place.	Time Class No. 2 One Hand 8″ = 0.011 One Hand 24″ = 0.013 Two Hands Sim. 24″ = 0.020  Dispose of toaster assembly to tray.
CONDITION C Positioning of parts on or into difficult or complicated locations. Assemblies or fixtures requiring the positioning of parts with respect to two definite points, or location in two directions.	Time Class No. 2 One Hand 8″ = 0.011 One Hand 24″ = 0.013 Two Hands Sim. 24″ = 0.020  Place screwdriver wrench over standard nut.	Time Class No. 3 One Hand 8″ = 0.019 One Hand 24″ = 0.021 Two Hands Sim. 24″ = 0.031  Place power screwdriver on head of self-tapping screw.
CONDITION D Positioning is much the same as Condition C but in addition may involve close tolerances, greater care of finishes, three or more points or directions of location, or application of force to assemble.	Time Class No. 3 One Hand 8″ = 0.019 One Hand 24″ = 0.021 Two Hands Sim. 24″ = 0.031  Place screwdriver on screw in assembly.	This combination not used.

Figure 25. Standard Time Values for "Place."
(Reprinted by permission from Barnes, *op. cit.*, opposite p. 345.)

SIZE OF OBJECT AND TYPE OF GRASP		PLACE TIME Class No.	STD. TIME Corrected for Transport Distances
(Small) 2 F Two fingers and thumb	(Very Large) 2 H Two hands		
Time Class No. 1 One Hand 8″ = 0.006 One Hand 12″ = 0.007 Two Hands Sim. 12″ = 0.011 Place small machine bolt in other hand. Rarely used other than for placing part into other hand.		1	0.007
	Time Class No. 2 Two Hands 8″ = 0.011 Two Hands 12″ = 0.013 Move waffle iron partial assembly to rough position under power screwdriver.	2	0.013
Time Class No. 2 One Hand 8″ = 0.011 One Hand 24″ = 0.013 Two Hands Sim. 24″ = 0.020 Place flat steel washer over stud or pin where tolerances are large.	Time Class No. 3 Two Hands 8″ = 0.019 Two Hands 24″ = 0.021  Place rectangular aluminum casting in fixture.	3	0.021
Time Class No. 3 One Hand 8″ = 0.019 One Hand 24″ = 0.021 Two Hands Sim. 24″ = 0.031  Position in two directions Place screw in tapped hole.		4	0.026
	Time Class No. 5 Two Hands 8″ = 0.030 Two Hands 24″ = 0.036  Place toaster fixture on locating pin for driving screw.	5	0.036
Time Class No. 4 One Hand 8″ = 0.024 One Hand 24″ = 0.026 Two Hands Sim. 24″ = 0.039  Place nut on terminal in limited space where fingers are cramped by design.	Time Class No. 6 Two Hands 8″ = 0.042 Two Hands 24″ = 0.048  Place unit cover plate over unit on waffle-iron grid assembly.	6	0.048

Corrected for Transport Distances. Time in Minutes.

stop watch. The procedure is based upon the assumption that any task can be divided into a number of component parts or elements, and that these elements will be identical with elements to be found in other tasks. If all jobs are regarded as made up of combinations of a limited number of basic movements, the standard time for the job is determined by adding the previously established standards for each of the elementary movements which compose it.

Figure 25, taken from Barnes,[22] illustrates the kind of data which are prepared for use in establishing synthetic time standards. Various conditions of placing a tool or part are classified, and the time for an element of placing which is involved in a task under study is read directly from the chart. After the times for all elements of the job are added, a fatigue and personal allowance is added to the total to give the standard time.

This method does not require any stop-watch study or rating in its application. Nevertheless, the rating is not actually eliminated from the complete procedure. The tables of standard times for each element were themselves originally established by the leveling method or some equivalent procedure. From this point of view, then, the only advantage for the synthetic method is that the rating is done once for all, and not all time-study men would need to be trained in rating. If the original tables of elemental times were worked out by a small number of highly trained raters, the inaccuracies due to unreliable ratings could thus be more easily reduced. Rating is still present, however, and no one knows how much more reliable the procedure might be.

In addition to the fact that the synthetic time method does not avoid the problems of leveling, it has certain additional drawbacks. While the errors of rating might be smaller with this system, what errors are made will be perpetuated, and will not vary at random from job to job as they would in the leveling procedure with stop-watch observations of each job standardized. The only hope is that errors in the standards for one element of a job will be canceled by opposite errors in other elements of the job.

The fundamental assumption used in synthetic methods is itself open to question. It is assumed that the time for a given motion should be the same, regardless of what other movements are included in the task. It would be just as logical to assume that the time for a given movement depends upon both the preceding and succeeding elements of the task. Such an assumption would, however, make the

---

[22] Ralph M. Barnes, *op. cit.* (opposite p. 345).

determination of synthetic times an impossibility. Thus the establishment of this method as an accurate device depends not only upon research into the accuracy of the leveling procedure, but also upon research into the additional assumptions required by the method.

One industrial engineer has carried this assumption even further by stating that the time for a simple movement is not only independent of other movements in the pattern, but that it is constant from individual to individual. As Segur states his principle, "Within practical limits, the time required for all experts to perform a true fundamental motion is a constant." [23] In demonstration, Segur predicts the time required for an individual to write his signature at top speed by analyzing the movements required.

The implication of Segur's "principle" would be that all persons doing a given task at maximum speed and with the same method (*i.e.,* exactly the same pattern of motions) would perform the task in the same time. In a letter to Gordy,[24] Segur states that the "practical limits" would be a maximum variation of 20 per cent. If this notion were accepted, the synthetic determination of standard times would be based upon times for "fundamental motions" (smaller subdivisions of the total activity than those illustrated in Barnes's figure on pages 230-231). The establishment of these times would require time studies under conditions of maximal effort or with a leveling factor for effort. No allowance would be made for skill, however, since the operator would have sufficient practice to become an "expert." The time for the total operation would be established for those movements which, in the opinion of the industrial engineer, make up the best method of doing the job. In practice, some account is taken of the combinations of movements involved in the job, since Pollock mentions the need for considering the overlapping of motions where the operator prepares for the next movement while finishing the current action.[25] In such cases, a combination-time is used, but evidently there could not be allowances for a great variety of interdependence of movements.

In support of Segur's method, Pollock, an employee of Segur, has presented several experiments in which the time for an operation was predicted according to Segur's method, and then the times required

[23] A. B. Segur, Labor costs at the lowest figure, *Manufacturing Industries,* 1927, *13,* 271.

[24] C. B. Gordy, Variations in the time required for skilled operators to perform a simple motor task, *American Journal of Psychology,* 1943, *56,* 181.

[25] K. W. Pollock, The use of therblig times for rate setting, *Advanced Management,* 1937, *2,* 35-40.

by a few operators were obtained by direct analysis of films.[26] In this rather small number of instances, the total times for the operations agreed closely with the predicted totals. The errors in the individual elements are much larger, however, and it would appear that the small error in the total is due to the random variation in these errors from element to element, so that most of the error is canceled out.

Gordy has put Segur's basic assumption to a direct test.[27] He carefully measured the times of over thirty operators performing a simple manual task requiring picking up and placing of blocks in a container. All the subjects were industrial employees in manual work, and all practiced the experimental task until they no longer improved before the times were recorded for analysis. They all used the same method or pattern of movements, so far as could be detected by the naked eye, although slight differences appeared when the motion pictures were studied. It would appear justifiable, therefore, to consider these subjects as "experts," and as using substantially the same movements. The results of Gordy's experiment are shown in Table 23.

TABLE 23*

MEAN CYCLE-TIME, STANDARD DEVIATION, AND RANGE
(*In "winks" or 1/2000th of a minute.*)

	Right Hand	Left Hand
Mean	835.12	882.97
Standard Deviation	87.54	96.81
Range of Individual Times	669 to 1133	702 to 1118

* *Ibid. ut supra,* 192.

It is evident that the range of times is such that the slowest operator requires about 70 per cent longer than the fastest. The extreme range could be more accurately estimated from the standard deviation multiplied by six. In both the left-hand and right-hand data, this would give a range of variation of over 30 per cent above and below the mean.[28] This is definitely greater than the range

[26] *Op. cit.*

[27] C. B. Gordy, *op. cit.,* 181-194.

[28] Gordy confuses the issue by calling attention to the standard error *of the mean.* This is the measure we would use if we wanted to predict the *mean* performance of another *group* of operators on the same task. Gordy states this fact correctly, but implies that it bears upon the validity of Segur's assumption. Segur's statement, however, has to do with individual variation, not with the variation of group means. In addition, the standard error of the mean could be reduced by merely increasing the number of operators in the experiment, thus attaining as small a standard error as is desired.

claimed by Segur. Even if we were to take twice the standard deviation as the practical range, thus bringing the result closer to Segur's claim, the data represent a very wide variation. We may therefore wonder what value a "practical limit" of 20 per cent might have in establishing fair time standards.

Since the subjects of Gordy's experiments were highly motivated and placed on their mettle by the tests, there can be little question that differences in skill are reflected in the speed of individual movements as well as in differences in method of work. Thus a leveling procedure must take account of both skill and effort or of some composite of the two factors. This point has bearing upon time-study method in general, as well as upon the specific procedure by which Segur established his tables of standard times for each element of motion. Pollock's data is not extensive enough to establish the validity of times estimated upon the basis of Segur's method, and Gordy's experiments show very conclusively that the method is based upon faulty assumptions.

This conclusion concerning Segur's "principle" does not, of course, invalidate all procedures of developing synthetic time standards. Even Segur's method might work, in practice, in spite of its faulty theoretical basis. In the last analysis, the value of any method of establishing synthetic times depends upon the validity of the time-tables which are used, and upon the validity of the fatigue or other allowances applied to the job. These basic data cannot be obtained until the fundamental problems of time study in general are solved. When these problems are solved with reasonable accuracy, synthetic methods may be a useful short cut, if it can be shown that the addition of elements with little regard to combinations and interrelations of movements is justifiable. Meanwhile the use of synthetic time procedures would appear to be a process of pyramiding methods of unknown validity, one upon another.

Summary.—In order to understand the operation of financial incentives, it is necessary to understand current practices which affect the accuracy and fairness of wage rates, particularly the *relative* pay of various workers. We have seen that time study plays an important role in establishing the financial incentives for many jobs. There is a great deal of mistrust of these methods on the part of the workers. While this mistrust is usually based upon false premises, and upon faulty understanding of the aims of time study, this mistrust is a potent factor in industrial motivation. Mathewson's

studies of restriction of output, mentioned in Chapter 8, bear this out.

If time-study methods are to be used at all, the mistrust of the workers must be overcome. The most satisfactory way of overcoming this mistrust would be to produce concrete proof that the methods are reliable and valid, since there could be little criticism of the expressed aims of time study. The mistrust is based upon the belief that time study fails to accomplish its aim of providing equal pay for equal effort, skill, and risk.

Unfortunately, time-study men have, in the past, shown little interest in providing definite statistical evidence of the effectiveness of their methods, and the methods have been accepted largely on faith. There is some indication that this condition is changing, but at present there is still little information available.

In the present chapter we have analyzed the assumptions underlying the various steps of time-study methods now in use. In doing so we have also tried to show what kind of research is necessary for evaluating these assumptions and their application in practice. On the basis of the large number of unchecked assumptions and procedures involved in current practices, it does not appear that a very convincing case can be made out for the accuracy of either relative or absolute standards of performance now in existence. Until research has provided a more certain basis for time standards, the effectiveness of financial incentives will continue to be variable and unpredictable.

It will be seen that the concepts of effort and efficiency of performance enter into this discussion in two ways. We began our discussion of incentives as a consideration of factors which affect the effort level, and consequently the efficiency, of performance. We now see that the establishment of adequate financial incentives must be based upon fundamental knowledge of effort, and upon adequate means of estimating or measuring effort.

Chapter 11

MERIT RATING AND JOB EVALUATION

We come now to two further procedures in industrial practice which affect the motivation of the worker. *Job evaluation,* as we have already seen in our discussion of time standards, is concerned with the determination of the basic wage to be paid for a given job. Time standards are supposed to tell us how much the "average" man should accomplish on the job. Average production on one job is worth more than average production on another, however, because one job may require more expensive training and education, requires the worker to take greater risks, or to work under more disagreeable conditions of heat, cold, or dust. *Job evaluation,* then, is the process of scaling jobs with respect to one another, taking into account all of the surrounding conditions and requirements of the job. It is an evaluation of the *job* rather than of men, since individual variations in productivity are held constant by basing the evaluation upon the standard man or norm for that work. Variations in pay, dependent upon individual performance, are then superposed upon the basic pay level derived from the job evaluation.

Merit rating serves a function similar to time standards or piecework prices in its effect upon financial incentives. In those jobs where objective production rates are inadequate for determining the value of a given worker in comparison to others on the same job, or where production rates cannot be easily measured, the value of the man is *estimated* by the supervisor. This estimation is the *merit rating,* also sometimes called the "efficiency" rating.

In many jobs, then, establishment of fair relative pay rates for men on the same kind of work depends upon the accuracy of merit rating, just as fair pay depends upon time-study methods in many other tasks. In both cases, the fairness of the monetary incentive also depends upon the proper relation between the basic levels of pay in different jobs and classes of work. If an assembler is paid less than a certain machine operator, when the assembler is certain that the *jobs* are of equal value, the effect is obvious. It is the task of job evaluation to make certain that such differences are justified.

As the last example indicates, it is necessary not only to have accurate and justifiable methods of attaining these ends, but also that they be convincing to the worker. The first step, however, is to determine the accuracy of the methods. If that is done, the process of convincing the worker would consist only in making clear to him the basis of the decision. If the methods themselves are unclear and of unknown accuracy, it is impossible to convince the worker that the results are justified, except by evasions and deceptions.

Job evaluation and *merit rating* are grouped together not only because of their bearing upon worker motivation in industry, but also because the basic techniques commonly employed for both of these purposes have many common elements. Both consist of ratings, or judgments, and in methods of combining judgments of a variety of factors into a single index. The only difference is that one involves rating of a *job;* the other, rating of a specific *worker.*

Merit rating is a factor in many other phases of motivation, as well as in establishing the pay rate of an individual. Since it is a general device for determining the value of the individual to the organization, it enters into decisions upon promotion, transfer, and layoffs, as well as pay-raises within a single job classification.

Merit Rating

Since merit rating is probably more familiar to most readers, the general problems of rating will be discussed first in their application to this field. Any case in which a foreman or manager decides that a given individual is of more value to the company than another involves a form of merit rating. In recent years this process has been systematized and standardized in the hope of securing more reliable and valid results. Most of these procedures consist of developing a standard rating-form which serves as a record of the foreman's judgment and is designed to aid him in making accurate judgments.

Various kinds of rating form have been used, but the one which is probably most widely known in industry is the one which we shall examine in some detail. Since most of the questions which arise with respect to this particular method are also applicable to the other methods as well, we shall lose little in generality by limiting our consideration to this one technique. The kind of rating which we shall examine in some detail is the *graphic rating method with a system of point scores.*

In this method, the designer of the rating-form decides that he wishes or needs to have the men rated upon a certain list of character-

PROGRESS RECORD

Name_____ Dept._____ Div._____ Date_____

Employee's Position_____ Job Class._____

Note This rating will represent in a systematic way your appraisal of the employee in terms of his ACTUAL PERFORMANCE ON HIS PRESENT JOB In the interests of furthering careful analysis, the following suggestions are offered regarding the use of this form

1. Consider only **one** factor at a time
2. Study each factor and the specifications for each grade
3. Review upon completion to see that the rating of each factor applies exclusively to the individual's ACTUAL PERFORMANCE ON HIS PRESENT JOB
4. Comment fully at bottom of page and on reverse side upon any matter which in your opinion needs explanation

PERFORMANCE FACTORS	PERFORMANCE GRADE				
	Far Exceeds requirements of this job	Exceeds requirements of this job	Meets requirements of this job	Partially Meets requirements of this job	Does Not Meet requirements of this job
QUALITY OF WORK Accuracy Economy of Materials Economy of Time (his own and others) Neatness Thoroughness	Consistently superior ☐	Sometimes superior ☐	Consistently satisfactory ☐	Usually acceptable ☐	Consistently unsatisfactory ☐
QUANTITY OF WORK Productive Output	Consistently exceeds requirements ☐	Frequently exceeds requirements ☐	Meets requirements ☐	Frequently below requirements ☐	Consistently below requirements ☐
DEPENDABILITY Follows Instructions Judgment Punctuality and Attendance Safety Habits	Consistently dependable ☐	Dependable in most respects ☐	Ordinarily dependable ☐	Frequently undependable ☐	Consistently undependable ☐
COMPATIBILITY Attitude Towards the Company Attitude Towards Supervision Co-operation with Fellow-Employees	Inspires others to work with and assist co-workers ☐	Quick to volunteer to work with and assist others ☐	Generally works well with and assists others ☐	Seldom works well with or assists others ☐	Does not work well with or assist others ☐

COMMENTS _____

Figure 26. Example of a Typical Merit Rating Form

istics—*e.g.*, "dependability," "quality of work," "knowledge of the job," and such. He then draws up a chart like that shown in Figure 26, upon which the foreman records the ratings in each trait for a given individual. The individual is then given a point score for each trait, and these points are added to give an over-all merit score which determines his pay or promotability.

Such a numerical score is presumed to be more accurate than traditional unsystematic and qualitative verbal descriptions such as "he is a very valuable man," "I would like to see him promoted to supervisor," and the like. The use of a merit rating-form is also believed to have other advantages besides that of increased accuracy. It provides a concrete record of the opinion of the supervisor from time to time. It requires him to consider all his men, and to consider a variety of qualities.

The rating can also be used for interviewing workers, explaining why they were not given raises, what additional training they should have, and so on. It is also suggested that it helps to train the supervisors to judge men.[1]

Reliability and Validity of the Graphic Rating Method.[2]—The value of merit rating stands or falls, however, upon its accuracy. "Objectivity" of the *record* is of no advantage if you are merely recording unreliable impressions. It does no good to explain to a worker that he was not promoted because of certain ratings if you have no evidence of the validity of those ratings.

We are asked to accept the rating-scale technique largely on faith, since there is no way of testing its validity directly. The rating scale is ordinarily used in situations where there is no other way of obtaining an estimate of the worker's characteristics except, possibly, for records of his production. Even where output records are available, it is apparently believed necessary to know more than the productivity of the worker, else the rating scale would not be proposed. As a result of the impossibility of direct validation, we are told what the rating scale is *expected* to do or what it *ought* to do, with little or no support for these claims.

In the following discussion we shall consider these supposed advantages of the graphic rating scale in some detail. We shall try to show that these claims are by no means easily acceptable, in fact, that some of them are logically contradictory. Throughout this discussion we shall assume that the merit-rating system is satisfactorily administered, with proper instruction and training for the raters. The difficulties we shall discuss are inherent in the technique itself, even when it is administered under the best possible conditions. How often these conditions are met is another question, although some

---

[1] J. Tiffin, *Industrial Psychology,* New York, Prentice-Hall, Inc., 1942, pp. 234-238, gives an excellent statement of these potential advantages of rating methods.

[2] This analysis of the accuracy of rating and the proposal for another approach appeared in slightly different form in the *Personnel Journal,* 1945, *24,* 6-15, under the title, "Merit rating criticized."

may admit that these rating scales are only too seldom used according to the plans of those who designed them.

Although it is not always explicitly stated, the graphic rating-form in such frequent use apparently aims to secure several advantages over unsystematized procedures which have been prevalent in the past and continue to be common. These expected advantages might be listed as follows:

1. Requiring the rater to consider a number of different "traits" or aspects of the man rather than a single total evaluation without any explicit analysis.

2. Defining these traits for the rater so that there is less likely to be disagreement upon the meaning of terms.

3. Emphasizing that individual differences involve a number of levels along a continuous scale of each quality. Thus it prevents such vague terms as "dependable," "skillful," "honest," being applied to a worker, and requires instead an estimate of the *degree* of dependability, skill, or honesty possessed by the individual.

4. By giving examples or definitions, it is hoped that the raters will agree not only upon the general meaning of the trait under consideration, but that they will also be able to use a common standard for determining the various levels of the trait. If this purpose is attained, "average skill" would mean about the same thing to various supervisors. (Not all graphic rating-forms make the attempt to define levels of a trait, and some even fail to define the traits themselves. We are considering here the more carefully worked-out rating techniques, and all of the possible advantages which might accrue through the use of all known precautions in composing the form. There are methods for "correcting" for errors due to failure to secure common standards of raters. These will be discussed later.)

5. By deciding upon the relative importance of the various traits considered in the rating, and by assigning point values to each level of each trait, it is possible to produce a total score which would represent the over-all value of a particular employee and would provide ready comparison with other employees.

It is not usually recognized that the first and last of these expectations are incompatible with each other. If we are successful in getting the rater to consider a number of distinct traits and phases of the worker, there are several difficulties and a definite fallacy involved in applying a total numerical score to the result. The fallacy consists in the assumption that it is possible to substitute one trait for another. The use of a total numerical score means that the worker

who is weak in skill can make up for that deficiency by getting an especially high rating in such a trait as dependability. Perhaps such compensation is possible within limits, but in cases where there is considerable divergence between the traits of a given individual, his total numerical score may not represent at all his relative value to the company. In the following example (Table 24) we shall for simplicity's sake suppose that each of three men is rated on three traits. They all receive equivalent total scores, but one would scarcely say that they are equivalent for purposes of transfer, promotion, or even for establishing relative pay.

TABLE 24

HYPOTHETICAL RATINGS OF THREE WORKERS WITH EQUIVALENT TOTAL SCORES

Trait	Point Values * of Ratings for Worker A	Worker B	Worker C
Skill	10	5	5
Cooperativeness	5	10	5
Capacity for future development	0	0	5
Total	15	15	15

* It is assumed that all three traits are of equal importance (*i.e.*, each is allotted the same number of points).

H. M. Johnson has pointed out this fallacy in connection with aptitude test-batteries which take into account several distinct aptitudes.[3] As an example, he describes the score which he might receive as an operatic tenor, based upon several component abilities. It is possible, he says, that his score for aptitude in this profession might be fairly good in spite of the fact that he is totally unable to produce a high note, and would therefore never have any chance of succeeding as a tenor.

Johnson's discussion was written primarily for psychologists. The point has received little consideration in discussions of merit rating which are aimed at the practical administrator, in spite of the probability that the fallacy is even more important in ratings than it is in the field of aptitude testing. In rating, we are not even limiting ourselves to aptitudes, but are considering all phases of the man, including a variety of traits which come under the broad heading of personality and character.

[3] H. M. Johnson, Some neglected principles of aptitude testing, *American Journal of Psychology*, 1935, *47*, 159-165.

Of course there are some arguments which could be advanced to justify the procedure of totaling points or scores based upon several distinct aptitudes, or even several disparate traits like those included in a merit rating-form. First, it might be asserted that if a particular trait is so important that the worker is no good without a high rating in that trait, the trait will be weighted in proportion to its importance. Thus, Johnson would not be able to get an extremely high score as an operatic tenor because ability to produce a high note would be weighted very heavily in the total score.

Unfortunately, there may be a number of traits, all equally important, and each of such nature that a low rating would indicate almost certain failure on the job. No system of additive point scores could take this situation into account. When we are dealing with aptitude tests, the process of validation would be such that these possibilities would be examined. If there were minimum requirements on each trait, these could be established through an examination of the validation data which would be secured in developing the test. It would be possible to discover *critical scores* which would be taken into account before any composite score is determined. These critical scores would appear in the correlation tables which are used in analyzing the validation data. If an additive procedure is finally adopted for combining the test scores, it can be based upon observed correlations with job performance, and the weighting based upon the closeness of relation between test performance and job performance. If the additive procedure is successful, it must mean that some degree of substitution of traits is possible—that the worker can make up for his deficiencies in one aptitude by having an extra amount of another.

These devices for overcoming the difficulty do not exist when we are dealing with merit rating. There is no way of weighting the various traits according to their importance, except by taking the consensus of opinion of those engaged in developing the rating scale. There is no way of validating these weights by a comparison with some independent measure, as there is when aptitude tests are properly validated. Since there is no independent measure of the value of the worker, there is no way of determining whether a low rating on a given trait is indicative of certain failure in the job or in other jobs to which the man might be promoted.

Thus, when we are discussing merit rating, it may be claimed that the system of weights attached to the various traits takes their relative importance into account. It may also be claimed that a worker can compensate to a certain extent for deficiencies in one trait by his higher degrees of other traits. These are only claims, however, un-

substantiated by any concrete evidence. In fact it is not possible to test these claims by any means known to the writer.

The logical difficulty which has been discussed here at some length seems to be the most fundamental difficulty with merit systems based upon a point system. If the first four of the advantages we have listed as possible attainments of a carefully designed rating procedure are secured in any degree, we should abandon the attempt to summarize the results in a numerical form. The only case in which the numerical summary could be justified would be that in which the raters do not actually rate distinct traits. Discussions of the "halo effect" have pointed out the tendency of raters to rate men similarly on all traits. If the halo effect is extremely strong, we might find that the rater does not really rate distinct traits at all. His separate ratings may be no more than repeated ratings of the worker on a single scale—a scale of the supervisor's belief concerning the worker's value to the organization.

If this were the situation, a numerical total score might be justified as being somewhat more reliable than the individual ratings. The advantage would be so doubtful, however, that we should seriously question the value of the rating technique. We lose most of the advantages which might be expected of the graphic rating method if the halo effect is strong, and the method no longer has much usefulness. Would it not be better to ask the supervisor to make a single careful rating of the over-all value of the man and let it go at that?

Ewart, Seashore, and Tiffin, in an analysis of a large number of ratings made in one industrial concern, found that there was a very strong halo effect.[4] It was so strong that ratings on one trait correlated with ratings on others to the extent of about .70, on the average. Tiffin suggests, however, that the result does not necessarily hold if the raters are given better training on the use of the rating-forms.[5] Bolanovich, analyzing the ratings of technical employees, found a somewhat lower average correlation.[6] Factor analysis showed six significant factors, while Tiffin found only three.

Thus we are faced with two possibilities. If the halo effect is as strong as Tiffin found it to be in his statistical study, few of the advantages which are claimed for graphic ratings are left. In such a case, no elaborate rating system is justified. If, on the other hand,

---

[4] E. Ewart, S. E. Seashore, and J. Tiffin, A factor analysis of an individual merit rating scale, *Journal of Applied Psychology,* 1941, *25,* 481-486.

[5] J. Tiffin, *op. cit.,* p. 245.

[6] D. J. Bolanovich, Statistical analysis of an industrial rating chart, *Journal of Applied Psychology,* 1946, *30,* 23-31.

the halo effect is only moderate in size, as it may be under careful training of raters, we may secure the advantages of a more complete analysis of the individual, and possibly some of the other benefits listed above under headings 1-4. We cannot in this case, however, justify the use of the total point score to summarize the results.

If the reader is still inclined to believe that substitution of traits is useful as a working hypothesis, in spite of the preceding discussion, we should like to point out other difficulties that still appear after this assumption is made. Another aim of the graphic rating method is to help the raters to adopt common standards of judgment. If this were fully successful, a man in one department who is rated at the "average" level in a certain trait would be equivalent to a man in another department rated at the same level. Students of rating data do not believe, however, that this condition is met in practice, since they usually provide methods of "correcting" for the raters' differences in standards.

We shall not here discuss the details of these techniques of correcting for constant errors of the rater. They may be found in numerous presentations of rating methods.[7] These methods, regardless of their details, all have a common basis. If one rater differs from another in the average ratings for all men in the department, it is assumed that this difference *must be* an error due to differences in standard between the supervisors. Before men from the two departments are compared, their ratings must be corrected for this constant difference between the raters. To put it another way, these methods assume that if the raters differ it must be due to a difference in leniency or kindheartedness. The possibility that the men in one department are actually better than those in the second department is not taken into account.

It is equally unjustified to make the opposite assumption that the difference of raters is due to a real difference in the quality of their men. No one knows which is the case, and the facts, if they could be found, would probably indicate that part of the difference is due to difference in leniency of the raters, and part is due to real differences between men in the two departments. Either the "corrected" or the "uncorrected" ratings, or both, may be in error by unknown amounts. What we have said here concerning differences in average scores is equally true of differences between raters in the *spread* of their ratings—another difference which is usually taken into account in "correcting" the ratings.

[7] For example, J. P. Guilford, *Psychometric Methods,* New York, McGraw-Hill Book Co., Inc., 1936, p. 274.

Thus, even if we are willing to make the assumptions necessary for using point scores for rating, there is little chance that we could with any degree of confidence compare ratings which are given by different supervisors. It is even questionable whether men rated by the same supervisor could be accurately compared if they happened to be working upon different jobs. The same rater's leniency may vary with the type of work, and we should be faced with the same dilemma we met in comparing results of different raters.

So far we have said nothing of the problem of reliability of the ratings. The direct study of the repeated ratings of a supervisor shows that they often have poor consistency, and the correlations between successive ratings by the same rater are moderate at best. As Tiffin points out,[8] the reliabilities are such that a point system of rating, based upon a range of four or five hundred points, is absurd because differences in the last place and even in the tens place are likely to be the result of chance variations rather than real variations between the persons rated. This difficulty with reliability could be overcome by using only five to seven rough groupings for the total score, rather than attempting to use a finer scale, and it might also be improved by better training of the raters. Improving the reliability of the ratings would not, however, overcome the objections to a point system which we have discussed in preceding paragraphs.

What is the solution of this problem? As we indicated at the beginning, there is a real need for intensive search for other modes of approach to the problem of evaluating employees. So far, these other methods are not available. Until they are, however, the writer believes that little is accomplished by elaborating and complicating the rating methods. Instead, he would recommend proceeding on a much simpler common-sense basis. The following suggestion is presented as one possible way out, not as a final solution of the problem.

A Suggested Procedure of Rating.—For lack of any better and more reasonable suggestion, we could assume that experienced supervisors ought to be able to distinguish, on an *absolute* basis, three levels of over-all value of men in their departments. If the men are engaged in quite distinct types of work, they would have to be divided according to occupation and rated separately and with reference to that occupation. In addition, it seems reasonable that the supervisors should be able to detect extreme degrees of certain specialized traits, even though they do exhibit the halo effect and unreliability when they are asked to rate all men in the department for each trait.

---

[8] J. Tiffin, *op. cit.*, p. 254.

The results of these suggestions would be a form like Fig. 27, with certain changes dictated by the industry or the occupations involved in a particular application. The form which is reproduced in Fig. 27 is not meant as a final version, and revisions of the instructions are probably necessary for use in any specific organization. The general approach is the thing which we are mainly interested in presenting at this time. It should be emphasized that this proposed form does not include anything radically new. In fact, it is basically a return to procedures of evaluation which are older and simpler than current rating techniques. It is an attempt to secure some of the advantages of a standardized rating technique without depending so heavily upon the assumptions required for the graphic scale-point method of evaluation.

INVENTORY OF PERSONNEL

Department _____
Jobs _____ (This should be a single job or group of similar, related jobs. Jobs requiring unrelated skills or training should be considered independently.)

I. All-round Value

1. *Outstanding Men:* List names of the men who are doing an exceptionally good job, and whose all-round value to the department is very high. Consider their present ability and skill, not what you think they might develop later. Men placed in this group should be in about the top 10 per cent of men in this line of work *anywhere,* and not only in this shop. If you are doubtful about any man's belonging in this group, *do not include him.* List these men in order from top down if that is possible.
2. *Poor Workers:* List the names of all men who are doing an exceptionally poor job, and whose all-round value to the department is very low. Consider present ability and skill, but do not include those who are poor because they do not have sufficient training or experience. These will be listed separately. As in the top group, if you are doubtful about a man's belonging in this group, do not include him. The men in this group should be those who would be considered as about the bottom 10 per cent of men in this occupation *anywhere,* but they may include more or less than 10 per cent of your own men.
3. *Average Workers:* List all men who have not been included in the first two groups, except those who are not fully trained or who lack experience. As you list them, consider each one so as to make sure that he does not belong either in Group 1 (exceptional men) or Group 2 (poor workers).
4. *Trainees and Inexperienced Workers:* Include here all those who have not finished the normal training period, or who have not worked long enough to show what they can do. Mark those who show promise with a *G,* and those who do not appear to make normal progress with a *P.*

Figure 27a. Suggested Form of Employee Appraisal (page 1)

II. Special Qualifications and Abilities

Listed at the left are a number of special characteristics which may be important considerations in making promotions or transfers. List, for each of these qualifications, the names of those men in the job you are considering who are outstanding in this quality, and those who are especially poor in this respect. The names listed are not necessarily the same names you have given in your over-all estimates of the value of the employee on page 1.

Quality *	Outstanding Men	Poor Workers
1. Dependability. (Does he follow instructions without close follow-up?)		
2. Ability to get along with others. (Is he respected and trusted by fellow workers?)		
3. Specialized technical knowledge of the job.		
4. Ability to instruct others in his job.		
5. Interest in his work and in the department.		

* Other traits to be added to meet the needs of a particular organization. Instructions to be revised after conferences with supervisors and experience in application of the forms.

Figure 27b. Inventory of Personnel (page 2)

The writer believes that this "Inventory of Personnel," though still having flaws which cannot be entirely eliminated, is more defensible than the graphic rating scale. It obtains an over-all estimate of the value of men in a job from the supervisor, which, in view of the halo effect, is all we can confidently expect of the graphic rating technique. It requires the supervisor to make only broad groupings, which, in the light of the known reliabilities of the graphic scale, are the only groupings justified. (If three groups are felt to be too broad, the same technique could be used with four or five categories instead of the three major groups illustrated in Figure 27.) The method obtains these broad groupings without resort to dubious point-rating scales.

The method also allows us to obtain useful information upon special

traits without forcing the supervisor to rate all workers on each trait. He can thus designate those who show noticeable excellence or deficiency, without being asked to make ratings on these traits where no information is available.

Last, but by no means least, this procedure should reduce the amount of "paper work." Though the supervisor has to write names instead of making check marks, he is not overwhelmed with a rating-form for each man. One form can serve for all the men on a job or a related group of jobs, so long as sufficient space is provided for listing the names. At the same time, the supervisor is required to consider each man because the form requires that each man must be listed somewhere.

Obviously, the inventory method here described cannot fit all requirements any more than can any other single procedure. It is felt, however, that it has advantages, both logical and practical, over the graphic rating method for a great many of the purposes for which the graphic method has been employed.

If it is still felt that the graphic form of recording supervisor's judgments is desirable, some of its difficulties could be avoided by other methods of handling the information it contains. The rating-forms might be sorted by inspection into three or four categories. With some training for those who do the sorting, it is possible that the method could be made quite reliable, and it would not be necessary to make the assumptions required by a point system of evaluation. The evaluators could take into account both the average level of the ratings for a given individual, and any marked deficiencies on single traits. In this way a rough qualitative analysis of the forms might produce a more adequate evaluation than would be possible by the point system.

Even this procedure would not, however, settle the question of whether to "correct" for the rater's tendencies to rate higher or lower than normal. That problem could never be settled unless there is some more objective criterion against which the ratings may be checked. If we had such a criterion, the ratings would no longer be necessary.

Conclusions on the Graphic Rating Method.—We have supported the thesis that the graphic rating scale with numerical scoring has serious logical difficulties and practical inadequacies. These flaws are so serious that they more than offset any advantages which are claimed for these rating techniques, especially since there is little more than common-sense argument to support these claims.

There is a serious need for research to develop new and fundamentally different methods of evaluating workers for promotion, transfer, or validation of employment techniques. Until this development is carried out, the writer has proposed to return to simpler procedures which require fewer doubtful assumptions. This simpler approach was illustrated by a suggested form for recording information in an "Inventory of Personnel." Such an inventory is aimed at securing some of the advantages of currently popular rating techniques, while making the task of the rater as simple as possible, and making fewer baseless assumptions.

Job Evaluation

Several points in the above discussion of merit rating apply almost without change to the problems of job rating. It is only necessary to make the minor changes required by the shift of emphasis to the job rather than the man. For those who are unfamiliar with the methods of job evaluation, however, it may be well to specify the method in a little more detail, and to show the general way in which the previous criticisms of merit rating would be rephrased in their application to job evaluation.

The process of job evaluation has, like merit rating, been carried out in unsystematic fashion for centuries. A similar trend toward increased systematization of the process and increased "objectivity" of method has developed in recent years. Again, there are variations in the detailed procedure employed, but the basic problems remain the same in the various specific methods.

The job evaluator must first decide upon a list of job characteristics which will furnish the basis for comparing jobs. He may, for example, use the list of factors given in Table 25, which shows the National Electrical Manufacturers Association's system of factors.[9]

The analyst then decides, upon the basis of judgment, the relative importance of these factors in rating the value of a job; (to be more exact, the value of the time of an average worker of average effort and working under average conditions). A scale of points is then assigned to each factor and a job rating scale drawn up in a manner similar to that used in merit rating. As a result of these steps, the number of points assigned to the factor of education, for example, may range from 0 to 40, while experience may be graded from 0 to

[9] C. H. Lawshe, Jr. and G. A. Satter, Studies in job evaluation: I. Factor analyses of point ratings for hourly paid jobs in three industrial plants, *Journal of Applied Psychology*, 1944, *28*, 189-198.

100, reflecting the judgment that the experience requirement is two and a half times as important as educational requirement in the value of a job.

TABLE 25

FACTORS IN JOB EVALUATION

Skill	*Responsibility*
Education	For Equipment or Process
Experience	For Material or Product
Initiative and Ingenuity	For Safety of Others
Effort	For Work of Others
Physical Demand	*Job Conditions*
Mental or Visual Demand	Working Conditions
	Unavoidable Hazards

Each job is then rated, and a total point score determined for it. Jobs may then be grouped into classes or labor grades, and to each grade is assigned a range of pay. Men in a given labor grade will then be paid upon the basis of this standard rate for the labor grade, with a correction dependent upon their merit rating, or bonuses for above-standard production. Job evaluation determines, therefore, the *relative* base pay for each job. The absolute amount of pay depends also, of course, upon the general economic conditions, the amount of profit alloted to labor costs, collective bargaining, and so on.

Let us look briefly at the parallels between problems of accuracy of job evaluation and of merit rating. At first glance, the relative weighting of job factors might seem to be open to the same doubts that were raised for individual traits. The methods of job evaluation assume that one factor can be substituted for another, just as merit rating assumes that one personal quality can be substituted for another. In job evaluation, however, there is probably better justification for this assumption. A job requiring no education but great risk certainly should be paid on a basis comparable to a job requiring much education but no risk. The problem of relative weighting is still insoluble except on the basis of opinion, but substitution of one trait for the other is conceivable and more readily justified in job evaluation.

The halo effect would appear at first to apply only to evaluations of a worker, and not to evaluations of a job. Nevertheless, there is something quite similar, if not identical, in job evaluation. This would be a tendency for a job which requires (for example) long

training to be rated high on other factors as a result of this one requirement. In fact, one statistical study has shown that the NEMA job ratings on specific factors are so closely intercorrelated that all of the variation in total point rating could be accounted for in terms of two or three independent factors.[10] Indeed, when only three factors were selected for rating, the ratings of the jobs were practically the same as they were in the more complicated rating based upon eleven factors.

As in the case of merit rating, the procedures are frequently used without check upon the reliability of the ratings. There is very little information available about the reliability of the method. When we turn to the question of validity, we find that job evaluation, like merit rating, cannot be validated. Both are estimates of something which cannot be measured and must be accepted on faith.

Except for a few points of difference, therefore, the parallel between the two applications of rating can be very closely drawn. As a result, our conclusions upon job evaluation will be very similar to those for merit rating. Job evaluation procedures should not be used without adequate analysis of the reliability of the results. The fineness of job grading must be in keeping with the reliability. Since validity cannot be established except by discussion of the basic assumptions made in the method, this is a matter for debate. Until workers are convinced that the various factors are given their proper relative weight, the procedures will not aid in providing adequate incentive conditions in industry.

Regardless of how we decide to gain common acceptance of the method as being valid, the evidence now available indicates that only a few factors should be considered. In fact, the whole procedure should be kept as simple as possible. Increasing complication not only adds nothing to the accuracy, but also makes the method more difficult for the worker to understand and accept.

---

[10] C. H. Lawshe, Jr. and G. A. Satter, *op. cit.*, and Studies in job evaluation: II. The adequacy of abbreviated point ratings for hourly paid jobs in three industrial plants, *Journal of Applied Psychology,* 1945, *29,* 177-184.

Chapter 12

PSYCHOLOGICAL PROBLEMS IN ACCIDENT CONTROL

It is obvious that accidents are inefficient, so much so that we should be willing to expend a great deal of extra effort and money in order to reduce the danger of accidents. It is also self-evident that many of the factors which increase the danger of accident are psychological as well as mechanical, chemical, or electrical. It might be said, in fact, that the control of the physical factors in accidents has now progressed far beyond our ability to control the psychological causes.

Although few will need to be convinced that accident control is an important problem, most of us have a very inadequate notion of just how enormous the problem is. In 1940, for example, there were about 96,000 deaths in the United States and about 9,000,000 injuries, all due to accidents. During the recent war, from December 7, 1941 to the surrender of Japan, our war casualties totaled 261,608 killed and 651,911 wounded, with 32,811 missing.[1] The total accidents in the United States during that same period resulted in 355,000 deaths, with 36,000,000 injuries. Table 26 shows how these totals are subdivided into various main groups.

TABLE 26*

TOTAL ACCIDENTS: DECEMBER 7, 1941 TO AUGUST 14, 1945

	Deaths	Injuries
All accidents	355,000	36,000,000
Motor vehicle accidents	94,000	3,300,000
Home accidents	118,000	17,500,000
Accidents to workers:		
Total	160,000	15,000,000
Occupational accidents	*66,000*	*6,600,000*
Off-the-job accidents	*94,000*	*8,400,000*

* From National Safety Council, Inc., *op. cit.*

General Psychological Problems in Accident Control.—It is, perhaps, within the realm of possibility that accident-*proof* machines and transport equipment could be designed and constructed. The

[1] National Safety Council, Inc., *Accident Facts,* 1946 edition, p. 17.

expense would probably be prohibitive, and even then we would not eliminate many of the accidents that are due to falls, bumps, and burns. Under present conditions, our machinery and equipment are less dangerous than they formerly were, and much care has been given to improvements in their safety characteristics. In the future, the greatest reductions in accident rates would appear to depend upon control of the individuals who have the accidents. Naturally, this broad problem has many subdivisions which must be considered independently of one another.

For convenience, we may divide the factors affecting accident rates into two main groups—general factors which affect most individuals in a similar fashion, and individual factors which make some individuals much more susceptible to accidents than others.

Propaganda and "educational" campaigns to reduce accidents would be included logically under the heading of general factors, since these campaigns are broadsides directed at the whole public or at workers in general. We shall make only brief comments about this propaganda factor, however. For one thing, these campaigns have grown up along with the development of various mechanical safety devices, improvements in factory design, methods of work, and all the changes in industry which have taken place in the past few decades. The separate effects of safety propaganda are therefore difficult to evaluate. There is little doubt that these campaigns, and the activities of such organizations as the National Safety Council, have had their share in the reduction of accident rates which has taken place during recent decades. It is the amount of their share which is an unknown quantity because of a lack of controlled investigation.

Propaganda for the reduction of accidents is essentially like commercial advertising or any other form of publicity. In order to improve it and make it still more effective than it is, it would be necessary to resort to the techniques of marketing research and methods of testing advertising effectiveness. Those who are especially interested in this problem should therefore look into the large literature upon these research techniques.[2]

Other general factors whose effects upon accident rate have been investigated include most of the specific factors governing efficiency which were discussed in Chapter 5—lighting, ventilation, hours of work, alcohol, and fatigue. Since the main outlines of these topics have already been presented, we need add only the special details related to their effect upon accident rate.

---

[2] *E.g.*, see H. C. Link, *The New Psychology of Advertising and Selling*, New York, The Macmillan Co., 1932.

1. Fatigue. The use of accident rates as a measure of the cost of work was taken up along with other indicators of fatigue and effort. This was not meant to imply, however, that we were assuming any direct correlation between effort, fatigue, and accident rate. Accidents themselves are costly, and as such must be entered in the ledger under the heading of cost of work or input. The relationship of accidents to other factors in cost of work still remains, however, to be discovered.

There is a common belief that accidents increase when the worker or driver is fatigued. There is no doubt that falling asleep is a definite and unequivocal cause of accidents. Falling asleep is not, however, a necessary result of fatigue, and it may occur as a result of other factors as well—we have little more than speculation upon the relationships involved. Falling asleep may be the result of bad ventilation, poor lighting, alcohol, carbon monoxide gas, boredom, poor motivation, and so on. Driver-asleep accidents do constitute a serious problem, and one which awaits further clarification by studies of factors in sleep as well as by direct analysis of this special situation.

The National Safety Council has made some statistical studies of the driver-asleep accident, although the collection of accurate statistics is extremely difficult.[3] As would be expected, most of these accidents occur during the hours between midnight and six in the morning. The peak hour for truck accidents occurs earlier than that for passenger cars. For passenger cars, the highest rate of accidents in which the driver is asleep occurs about eighteen or twenty hours after the last sleep of the driver. Among truck operators there is also a high rate of this kind of accident, about five hours after the last sleep. The authors of this study suggest that the truck drivers obtained so little sleep that they became drowsy more quickly.

An estimate of the absolute frequency of accidents involving sleeping drivers is almost impossible to make because even if the driver lives through an accident he is reluctant to admit that he was asleep. There are enough substantiated cases, however, to indicate that the problem is serious. The National Safety Council Report also includes data secured by stopping trucks on the highway and asking them when they had last slept and how much. As many as 10 per cent of the drivers thus questioned had been on a trip without rest as long as twelve hours, and the same proportion had been awake as long as eighteen hours.[4]

[3] National Safety Council, Inc., *How Long on the Highway?* 1937.
[4] *Ibid.*, p. 22.

The problem of permissible driving hours for vehicle operators has been attacked by means of fatigue tests. We have already illustrated this mode of approach in our general discussion of fatigue tests where we referred to the example of Ryan and Warner. (See pages 82-86.) The difficulties of interpreting the results of such a research were outlined in some detail on those earlier pages. While the findings include significant changes in a variety of functions as a result of a long period of driving, there is no way of determining when the fatigue is serious enough to make driving hazardous.

A somewhat similar approach was made by a large group of investigators whose work was sponsored by the Interstate Commerce Commission.[5] This research has been severely criticized by Johnson, not only on the grounds which make most fatigue-testing programs of doubtful practical value, but also for additional special features of the technique employed in this particular program.[6] Instead of testing the *same* drivers before and after driving, and also on control days when they did not drive, the investigators resorted to mass testing of drivers (usually a single battery of tests for a driver). Then the drivers were classified into groups according to the number of hours they had driven since their last sleep, and the groups were compared for differences in performance on the tests. This procedure obviously introduces many additional uncontrolled variables which make it difficult to obtain significant results. Differences between groups would be affected by the different individuals included in them, although this could be controlled by large numbers if there was no systematic error of selection in the various groups. In addition, as Johnson points out, the norm with which each score was compared was based upon drivers who had not *driven* since their last sleep, the effect of their other activities not being taken into account.

Finding that four tests showed significant changes with increasing hours of driving, the authors then proceeded to work out a composite *index* of fatigue based upon these tests (speed of tapping, reaction co-ordination time, simple reaction time, and hand steadiness). Against this procedure Johnson raises the same objections which we have discussed earlier in merit ratings, where a composite score is worked out for a number of disparate qualities of the worker. Such a composite is, according to Johnson, a meaningless number which

---

[5] B. F. Jones, R. H. Flinn, and E. C. Hammond (together with other collaborators), Fatigue and hours of service of interstate truck drivers, U. S. Public Health Service, *Public Health Bulletin No. 265* (1941), xxiii and 286.

[6] H. M. Johnson, Index numerology and measures of impairment, *American Journal of Psychology,* 1943, *56,* 551-558.

contributes nothing to the solution of the problem of permissible hours of work for truck drivers.

Regardless of whether Johnson's criticism of the composite score is justified, the value of such a procedure of testing with or without a composite score is questionable for the reasons we have outlined earlier. It might also be mentioned that those differences in test score which were found to be significant were not large in a practical sense. They are significant because of the large numbers tested, but there is much overlapping between groups. The overlapping may be due to the poorly controlled research technique, or it may be that these tests are affected in a variable manner by driving, regardless of the adequacy of control.

It should also be said that a variety of medical tests were included in the program, mostly with negative results. The blood count of white corpuscles showed a significant change, with the highest value for the group who had driven an intermediate number of hours. None of the findings, however, made it possible to set limits to driving hours upon medical grounds.

Johnson suggests that a more useful first step in investigating this general problem would be to collect records showing the number of hours driven by each person involved in an accident. With proper statistical treatment, and with adequate numbers in the samples involved, it should be possible to determine the effect of length of work upon liability to accident. Since this is, after all, the main problem in accident control, it would appear more sensible and more direct to make this approach first.

Fatigue and Industrial Accidents. In seeking evidence as to the influence of fatigue upon industrial accident rate, most investigators have resorted to a study of hourly accident rates. If accidents increase in frequency as the day progresses, the assumption would be that fatigue is responsible for the accidents. A rough hourly pattern of this kind has been frequently observed (Figure 28), and the assumption of the influence of fatigue has followed the discovery.

Vernon, however, has raised strong doubt as to the validity of this assumption.[7] In several extensive surveys in munitions plants during the first World War he was able to find no marked increase in the accident rate in the afternoon as compared with the morning rate. (See Table 27.) He points out that variations which do occur from

[7] From H. M. Vernon, *Accidents and Their Prevention.* By permission of Cambridge University Press (England) and The Macmillan Company, New York.

hour to hour are closely related to the rate of work. The peak accident rate came in the next to the last hour of the morning, which was also the period of highest hourly production. This was the pattern for day shifts. In night shifts the pattern tended to be much differ-

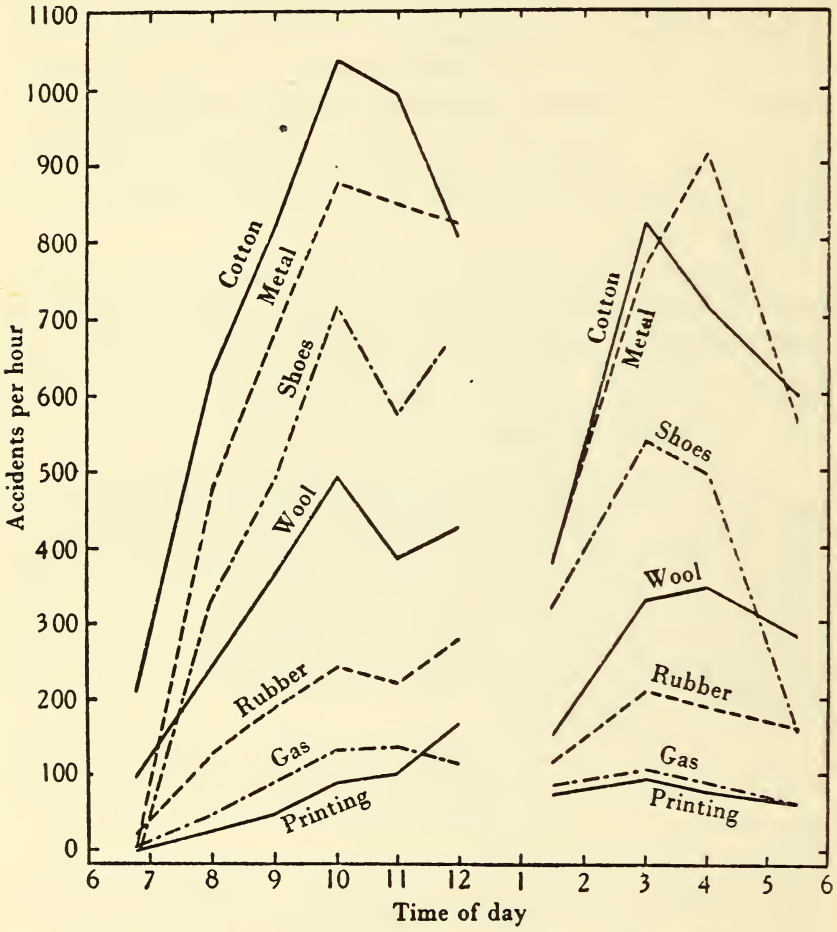


Figure 28. Hourly Incidence of Accidents in Various Industries
(From Vernon, *op. cit.*, p. 66. By permission of The Macmillan Co., New York.)

ent, with the maximum accident rate at the beginning of the shift and a progressive decline from that time throughout the working hours. At night, therefore, the factors differ from those that obtain during the daytime shifts. Vernon believes that the difference lies in the fact that the night-shift workers have been active with outside affairs before coming to work, thus tending to be distracted by these affairs

and consequently in a lively state. During the day shifts, "liveliness" tended to follow the same pattern as the output rate.

TABLE 27 *

Comparison of Morning and Afternoon Accidents

	Relationship of Afternoon to Morning Accidents
10 industries running 57.1 hours per week	96 to 100
8 industries running 49.9 hours per week	108 to 100
Mean: 18 industries running 53.9 hours per week	101 to 100

* Vernon, *op. cit.*, p. 67.

Vernon did find, however, some evidence of an influence of fatigue. When the working day was exceedingly long (twelve hours), the accident rate for women was much higher on an hourly basis than it was for a ten-hour day. It was also noted that the afternoon accident rate for women was 45 per cent higher than the morning rate in the twelve-hour day. (See Table 28.) Increased length of the working day did not have the same effect upon men, however. Their accident rate remained about the same on an hourly basis, and there was no increase in the ratio of morning mishaps to afternoon accidents.

TABLE 28 †

Effect of Length of Working Day Upon Increase of Accidents in the Afternoon

Length of Work Day	Statistical Period	Total Cuts Treated		Relative Accident Frequency per Shift		Percentage of Afternoon Accidents as Compared to Morning Accidents	
		Men	Women	Men	Women	Men	Women
12 hours	November, 1915 to January 1916	660	1145	114	273	107	145
10 hours	February, 1916 to March, 1917	2304	2426	100	100	108	117

† *Ibid.*, p. 68.

The greater susceptibility of women workers to long hours was also reflected in the records of the first-aid room in the factories under

study. There was a marked increase in cases of faintness in the twelve-hour day, as compared with the ten-hour period. Vernon reports this change on a relative basis—the ratio of women's cases to those of the men changing from 3 to 1 to 9 to 1.

Vernon's careful analysis would therefore indicate that progressive fatigue during the working day is not an important factor in accidents for our present industrial labor standards. At least the fatigue factor is not clearly evident in the over-all statistics of a whole factory or industry. If statistics for certain dangerous and exacting occupations could be collected in sufficient numbers, fatigue might be found to have a more marked and serious effect. In addition, it is possible that the effects of fatigue are disguised by the more powerful factors of rate of work and what Vernon calls "liveliness." If these factors could be studied separately and allowed for in the analysis of fatigue as a factor, our conclusions could be more definite.

After Vernon first made this analysis, the United States Public Health Service made an analysis of accidents in which the effect of production rate was held constant.[8] The method was to divide the accident index by the production index for a given period; in other words, to report in terms of *accidents per unit of output.* In the earlier hours of the day, this accident index rises and falls with the output rate. (See Figure 29.) That is, production increases bring with them a disproportionate increase in accidents, as noted by Vernon. In the closing hours of a ten-hour day, however, this relationship breaks down, and the accident index remains high relative to production rate. Thus, by more detailed analysis it is possible to show an effect of fatigue upon accident rate, once the more potent factor of production rate is controlled.

2. Alcohol. The effect of drinking upon traffic accidents is difficult to isolate, since the driver who has imbibed is likely to be suffering from lack of sleep, unusual meals, or the excitement of a party or date. The tabulation of accidents in which drivers who had been drinking were involved will not indicate the importance of alcohol *per se,* therefore, but the total result of all these factors associated with drinking. This is our only source of information upon the problem, however, since controlled experimentation is obviously

---

[8] J. Goldmark, M. D. Hopkins, P. S. Florence, and F. S. Lee, Studies in industrial physiology: Fatigue in relation to working capacity, 1. Comparison of an eight-hour plant and a ten-hour plant, U. S. Public Health Service, *Public Health Bulletin No. 106,* 1920. Although this paper appeared soon after Vernon first published his analyses, the latter takes little account of these findings in his more recent book, already cited.

out of the question. Laboratory studies are sometimes made the basis of conclusions concerning the dangers of driving when under the influence of alcohol, but we have already pointed out that the laboratory tests do not necessarily correlate with the competence of a subject who is performing a different task in operating a machine or vehicle.

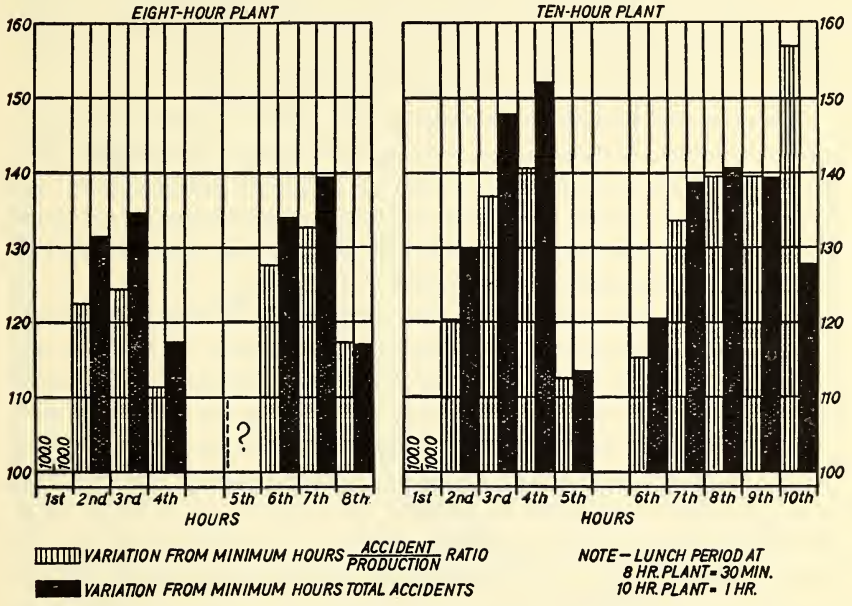


Figure 29. Accidents and Production in Eight-hour and Ten-hour Metalworking Plants

Eight-hour plant data include one year for three shifts, ten-hour plant data, two years for one shift.

(From Goldmark, *et. al., op. cit.*, p. 115.)

The central problem concerns those who have taken relatively small amounts of alcohol that are insufficient for definite intoxication. Everyone would recognize the danger of allowing an inebriated person to drive. The problem is how much alcohol will produce a significant increase in liability to accidents.

From the scanty data which are available, it would appear that even small amounts of alcohol increase the average accident rate. Vernon reports experiments in which driving skill was tested in an apparatus which closely resembled the actual performance of driving.[9] After imbibing two or three ounces of whisky, the subjects

[9] H. M. Vernon, *op. cit.*, pp. 159, 166.

increased their speed of driving but decreased their accuracy of control. Usually they were not aware that they had speeded up. More direct evidence of the influence of alcohol upon accident liability is afforded by the data of Holcomb.[10] He tested the alcoholic content of the urine of drivers injured in accidents. The percentage of these drivers whose tests showed that they had been drinking was several times as great as that among a random group who had not been involved in accidents. The proportion of injured drivers whose tests were indicative of intoxicating amounts of alcohol was thirty-three times as great as that of the non-accident group.

There is still less direct evidence of the effect of alcohol in industrial accidents. Vernon cites over-all reductions in accident rates which followed wartime restrictions on alcoholic consumption, and hours of work which prevented drinking. Such correlations must be treated with caution, however, since there are so many other factors which might make them spurious.

As Johnson has pointed out, the relationship of accident liability to alcoholic consumption is likely to be very complex.[11] Johnson was referring especially to the tendency toward compensation. It is even possible that a driver could *improve* his driving performance after drinking, because he knows the danger and devotes himself to his driving with extra care and attention. Jellinek and McFarland have concluded that the effects of alcohol are smaller for habitual tasks than they are for the kind of task used in the laboratory or for any unpracticed activity.[12] This possibility makes it all the more difficult to apply the laboratory results of alcohol experiments to road conditions. These authors refer to an experiment of Bauer's in which real driving reactions were tested after alcoholic dosage. Five out of the eight subjects showed a *decreased* reaction time in this test. If we accept Jellinek and McFarland's conclusion, the usual driving reactions might not be much affected by moderate consumption of alcohol. At the same time, however, we must consider the possibility that many situations which produce accidents involve a sudden emergency that calls for an unusual and relatively unpracticed response.

The only way in which such questions can be settled, therefore, is by direct study of the relations between the ingestion of alcohol and actual accidents, with suitable controls to eliminate the influence

---

[10] R. L. Holcomb, Alcohol in relation to traffic accidents, *Journal of the American Medical Association,* 1938, *111,* 1076-1085.

[11] H. M. Johnson, Pre-experimental assumptions as determiners of experimental results, *Psychological Review,* 1940, *47,* 338-346.

[12] E. M. Jellinek and R. A. McFarland, Analysis of psychological experiments on the effects of alcohol, *Quarterly Journal of Studies of Alcohol,* 1940, *1,* 272-371.

of other independent factors. What data we have indicate a steady increase in liability to driving accidents correlated with increases in the alcoholic content of the blood. Whether the same relationship holds for industrial accidents remains to be determined.

3. Lighting and Temperature. The effect of poor visibility upon accident rates is so obvious that it requires little analysis. This does not mean, however, that the problem is simple. There is still

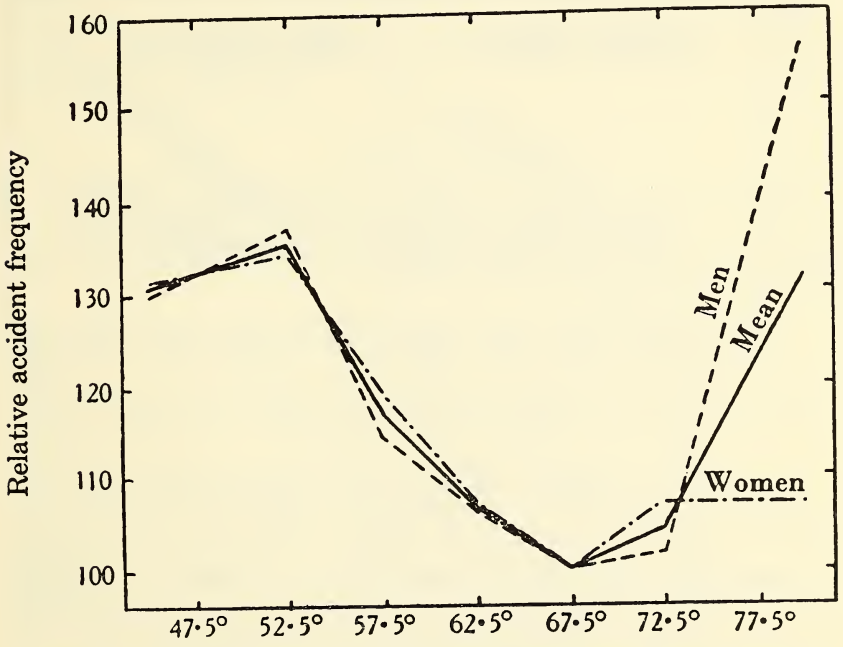


Figure 30. Accident Frequency in Relation to Temperature (Shell Factory)
(From Vernon, *op. cit.*, p. 76.)

much to be learned about highway lighting, headlights, visibility of signs and pedestrians, adequate lighting of machinery, and so on. Our previous discussion of the general problems of efficient lighting have shown the complexities involved. To detail the problems of vision in relation to accidents in industry and on the highway would require a separate treatise of a size out of proportion to the factual information available.

Temperature has an effect by increasing the hazards of the road, but it also has an effect upon the worker himself which sometimes increases his accident liability. Variation in temperature appears to affect the rate of minor accidents more markedly than the rate of

severe accidents. Vernon and his collaborators found that the incidence of cuts in a shell factory was at a minimum for a temperature of 68° to 70°, with increases in frequency as the temperature fell below or rose above this range [13] as shown in Figure 30. In coal mines, high temperatures increased the over-all rate of accidents and the amount of lost time, but the change was most noticeable in those accidents which resulted in less than ten days absence from work.[14]

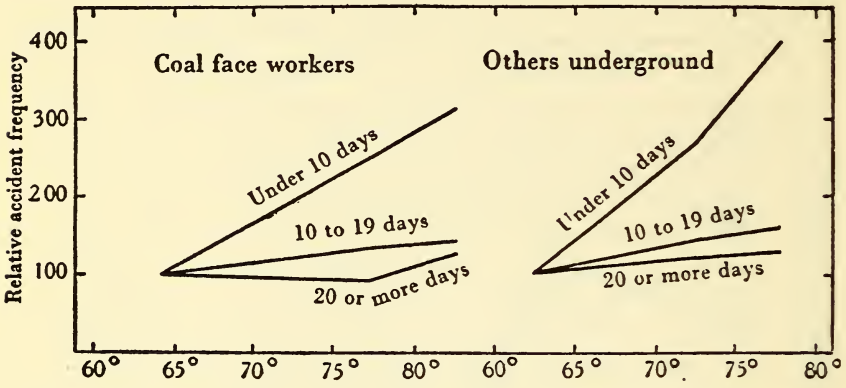


Figure 31. Accident Disability in Relation to Temperature (Coal Mining)
(From Vernon, *op. cit.*, p. 80.)

We are still unable to explain these relationships. The lower temperatures may produce clumsiness and stiffness which would account for the increase in minor injuries. The effect of the higher temperatures is more difficult to account for. It might be suggested that the high temperatures serve as distractions, but it is also possible that the change in temperature produces some change in ease of coordination, or increases fatigue.

Individual Differences in Liability to Accident

Although it is commonly recognized that some persons tend to have more accidents than others, the full implications of this fact are not always realized. Part of the difference in accident liability is due to conditions which are out of the control of the individual

[13] E. E. Osborne and H. M. Vernon, Two contributions to the study of accident causation: A. The influence of temperature and other conditions on the frequency of industrial accidents, Industrial Fatigue Research Board (Great Britain), 1922, Report No. 19, 1-17. H. M. Vernon, T. Bedford, and C. G. Warner, A study of absenteeism in a group of ten collieries, Industrial Fatigue Research Board (Great Britain), 1928, Report No. 51.

[14] Vernon, *op. cit.*

worker or driver—different jobs involve different amounts of risk, and some automobile drivers expose themselves to greater risks because they must drive farther than others or under poor driving conditions. Table 29, containing data collected by Tiffin in a steel mill, shows a wide variation in frequency of mishaps which occur in different jobs. Information of this kind shows the safety engineer where his more serious problems lie. For the psychologist who wishes to study individual differences in tendency toward accidents, differences in exposure and job hazard represent variables which must be held constant before the effects of the individual's characteristics can be observed.

TABLE 29*

VARIATION IN THE AVERAGE NUMBER OF HOSPITAL VISITS PER YEAR AMONG EMPLOYEES ON ELEVEN DIFFERENT JOBS

Job	Average Number of Hospital Visits per Year
Craneman	3.55
Opener	3.54
Reckoner	2.96
Machinist Helper	2.77
Leader	2.75
Sheet Inspector	2.54
Shear Helper	2.40
Assorter	2.36
Potman	2.10
Foreman	1.16
Roll Turner	.47

* From J. Tiffin, *Industrial Psychology*, New York, Prentice-Hall, Inc., 1942, p. 288.

Even within a group of workers who have equal hazards in their jobs, there will be a consistent tendency for certain workers to suffer more accidents than others. Those workers who have a high accident rate in one year will also have a high average rate in succeeding years. Such a worker, with a permanently high liability to accident, is spoken of as *accident-prone.*

The psychological problem is to find the causes of such accident-proneness. It is insufficient to attribute the accident tendency to "carelessness," "clumsiness," or "faulty attitude," as has sometimes been done. These terms mean very little, and do not solve the problem. We still need to know *why* the individual is careless or has faulty attitudes.

There has been a considerable number of exploratory researches aimed at finding underlying factors responsible for accident-proneness. The basic procedure of these researches has been very simple. A group of accident-repeaters and a group of accident-free workers or drivers are given a variety of psychological tests, such as reaction time, judgment of speed of movement and distance, muscular coordination, and so on. The average score of each group on each test is computed in order to see if the repeater group is significantly poorer at the test than the accident-free group.

There are several serious difficulties with such a gross attack upon the problem. First of all, the group of accident-*repeaters* is probably not made up entirely of accident-*prone* individuals. Notice that the definition of accident-proneness is in terms of *liability* to accident, *i.e.,* in terms of the probability that the individual will have an accident within a certain period of time.[15] It is not stated in terms of the actual accident record of the individual. A person who has a large number of accidents is not necessarily accident-prone. He is likely to be accident-prone, but he may be, instead, a person who has had bad luck—who has suffered accidents in spite of a low probability of accident. We can illustrate what is meant here by a reference to games of chance. If I toss a coin five times and obtain five heads, the coin is not necessarily loaded. Even if I repeat the trials and again obtain five heads it would not be conclusive proof of loading. I might strongly suspect loading, but it is possible to obtain this result by chance with a fair coin. Most of the repeater group is probably accident-prone if the group is properly selected. Some, however, are not accident-prone, and we have no way of telling in advance which individuals these are.[16]

---

[15] To be more exact, liability is defined as a limit. It is the average number of accidents per year (month, etc.) which an individual would have if he were exposed to the same conditions for an infinitely long period of time. Defined in this way, liability of an individual can never be measured, but only estimated. It is like the "true" score of an individual on an aptitude test, the score which he would attain if we had a perfectly reliable test. Accident rate over a period of a few years is, however, less reliable as an estimate of the true score, than most aptitude tests.

In spite of this difficulty we need the concept of accident liability. Accident rate cannot be predicted as such because it is determined by many truly accidental factors (luck or chance) as well as by permanent tendencies of the individual, and only these permanent tendencies are predictable.

[16] Essentially the same point of view is expressed by Cobb in relation to *correlations* between test scores and accident records. P. W. Cobb, The limits of usefulness of accident rate as a measure of accident-proneness, *Journal of Applied Psychology,* 1940, *24,* 154-159. For a more complete mathematical treatment of the problems of accident distribution, see E. M. Newbold, Practical applications of the statistics of repeated events, particularly to industrial accidents, *Journal of the Royal Statistical Society,* 1927, *90,* 487-547.

Similarly, the comparison group of non-accident individuals probably contains some accident-prone individuals. An accident-prone individual may have very good luck and thus avoid accidents in spite of his high liability. The result is that we are comparing the test results of two mixed groups. A difference which might exist between truly accident-prone individuals and persons who are truly low in accident liability may therefore be partially covered up or reduced by the mixture of the two groups. This is especially likely when the accident histories of the experimental groups are known over only a relatively short time. In the long run, liability and accident record tend to agree, and therefore we must have a long-term history of the individual, and he must be exposed to similar hazards over that long history if we are to get reliable results.

In order to make the above ideas more concrete, Table 30 shows an example of the distribution of accidents among a large group of automobile drivers. These data show the characteristics which have been found in a number of similar statistical investigations. The column headed "Actual Number" shows the number of drivers who experience 0, 1, 2, and so on accidents during the period of the survey. The column labeled "Expected Number" represents the distribution which would be expected if all the drivers had the same liability to accident (a liability equal to the average liability of the actual group). The difference between the "expected" and the actual distributions indicates that liability must differ from person to person, as we have already indicated. It also illustrates the point we have made with respect to the procedure of validating tests for accident-proneness.

In order that the statistical relations may be brought out, an additional distribution of accident frequencies is given in Table 31. This distribution is a hypothetical set of data chosen for ease of computation, but it will be seen that the general shape of the distribution is essentially the same as that given in the real distribution of Table 30. This hypothetical distribution was obtained by assuming that there were two separate groups of drivers or workers. One group of a hundred individuals averages two accidents per person over a period of observation. The larger group of a thousand individuals averages only 0.5 accidents per individual—only one fourth the rate of the accident-prone group. Adding the two distributions together resulted in the distribution which is labeled the "observed" distribution. The "expected" column is based upon the assumption of equal liability for the whole group of eleven hundred individuals who have an average of .64 accidents per individual. The divergences between the

"expected" and "observed" values are seen to be like those in the actual case of Table 30.

TABLE 30*

DISTRIBUTION OF AUTOMOBILE ACCIDENTS IN A RANDOM SAMPLE
(CONNECTICUT, 1931-36)

Accidents per operator during experience	Operators having these accidents		
	Actual Number	Expected Number	Difference
0	23,881	23,234	647
1	4,503	5,572	— 1,069
2	936	668	268
Total (0-2)	29,320	29,474	— 154
3	160	53	
4	33	4	
5	14		
6	3		
7	1		
Total (3-7)	211	57	154
Total	29,531	29,531	0

* U. S. Bureau of Public Roads, *Motor Vehicle Traffic Conditions in the United States: The Accident-Prone Driver.* (House Document No. 462, Part 6, 75th Congress, Third Session), 1938 (Washington), p. 19.

In the last two columns of the table are given the distributions of accidents in the two hypothetical groups which make up the total. It will be seen that a substantial number of the accident-prone individuals have zero to one accident apiece. Also, the low-liability group contains some individuals with three or four accidents. If we were to choose for study those accident "repeaters" who have three or more accidents, we should include thirty-two accident-prone individuals and fifteen low-liability individuals. The zero accident group would include fourteen individuals whose liability is high and should be called "accident-prone."

It is obvious that this example oversimplifies the problem. In real groups of individuals there are likely to be many different levels of accident liability, and the total accident distribution must be conceived as the sum of all these different groups. The same general principle of variation in accident *rate* for persons of equivalent *liability* will still hold, however, so that we have done no violence to the principle in choosing a simpler case for illustration.

TABLE 31 *

Hypothetical Analysis of Accident Distribution with Two Liability Groups

Number of Accidents	"Observed" Number of Individuals	"Expected" (Assuming equal liability)	Accident-prone Group * (2 accidents per person)	Low-liability Group * (.5 accident per person)
0	621	584	14	607
1	330	370	27	303
2	102	117	27	75
3	31	24	18	13
4	11	4	9	2
5	4	1	4	0
6	1	0	1	0
7	0	0	0	0
Total	1,100	1,100	100	1,000

* The "expected", hypothetical accident-prone and low-liability distributions are estimated from tables of the Poisson distribution. (H. F. Dodge and H. G. Romig, Single sampling and double sampling inspection tables, *Bell System Technical Journal*, 1941, *20*, Figure 6, opposite p. 20.)

There is still another condition which tends to vitiate results which are based upon the gross procedure of comparing test results of two groups of individuals of different accident histories. This is the obvious fact that there are many different causes of accident-proneness. Even if we are able to sort out a reasonably "pure" group of accident-prone individuals, some of the group may be prone because of various ocular deficiencies, some because of poor muscular coordination, others because of certain faulty habits of machine operation which have not been corrected, and so on. Suppose, now, that we wish to find out whether a specific factor, such as visual acuity, is a factor in accident liability. If we follow the gross procedure which we have outlined, we administer a test of visual acuity to our two groups of workers or automobile drivers. Suppose, also, that there is a certain small group of workers who are accident-prone because of defective visual acuity, and for no other reason. This group will probably be at most a small fraction of the total group of accident-prone individuals. The remainder may show normal or even superior visual acuity. The net difference in visual acuity between the two groups is therefore likely to be so small that it is scarcely detectable, even though visual acuity is (hypothetically) a definite cause of accidents in certain individuals. When we consider all the factors which might conceivably make an individual liable to accident in some complex task, it is evident that no single test could be ex-

pected to show very large differences between groups selected in terms of gross accident records alone. Negative findings for any particular test may mean either that the test performance is completely unrelated to accident liability, or that the factor tested is important in such a small group that its effect is hidden.[17]

It might appear that both the difficulties in validating tests of accident-proneness which we have discussed apply only to the particular kind of analysis which we have described. That is, they might appear to be a result of the process of selecting an accident-repeater group and an accident-free group and comparing average test scores for the two groups. Essentially the same problems arise, however, in any procedure for relating scores on a single test to gross accident liability. Liability is never directly measured; it is estimated from the accident record of the individual. The only way this estimate can be improved is by using longer and longer periods of observation. Gross accident liability, in turn, is a function of many variables, so that no one test can show extremely high correlation with liability, to say nothing of accident history.

Later we shall make some suggestions for new kinds of analysis which should establish more meaningful relations between tested characteristics of an individual and his accident tendencies. Before we do this, however, it will be necessary to become acquainted in a general way with the results now available—with the kinds of test which have been employed, and with the character of the statistical findings. In view of the logical difficulties we have described, it is surprising that so many significant relationships do appear in researches based upon the methods we have described.

Some Tests of Accident-Proneness.—The more frequently used tests may be grouped roughly as follows:

1. Tests of reaction time and co-ordination (frequently in the form of a miniature of the car controls or machine).
2. Tests of special visual characteristics (judgment of speed, distance, sensitivity to glare, peripheral and central acuity).
3. Distractibility and reaction in an emergency.
4. Intelligence.

1. Reactions and Co-ordination. As selected examples of this kind of test we may refer to those used by DeSilva and to those of Farmer and Chambers. DeSilva, directing an extensive program

[17] See Farmer's discussion of this problem in his comments appended to Miss Newbold's statistical papers. (*op. cit.*)

of research upon automobile drivers, made use of a test consisting of "dummy" automobile controls (Figure 32) which the subject manipulated while looking at a screen upon which appeared signals and an image of a highway scene.[18] The subject was tested for speed of response to a traffic signal (the time from the signal to the first

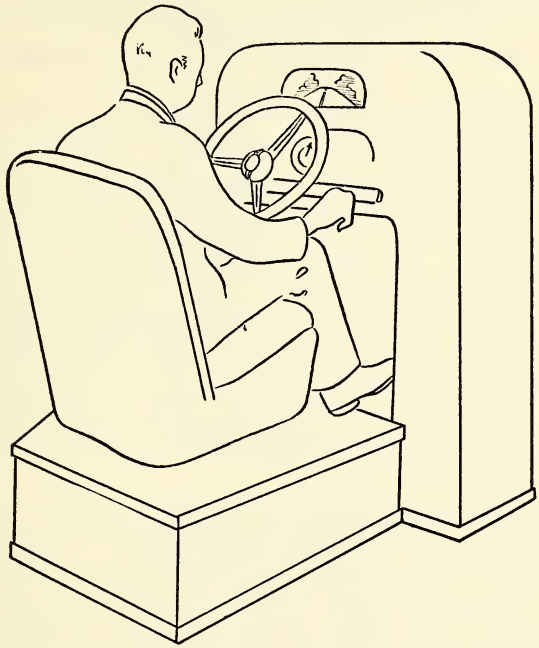


Figure 32. Driver Testing Apparatus with Moving Road Scene
Used for Harvard Tests of Braking Reaction Time, Steering, and "Vigilance Reaction."

(Reprinted by permission from *Why We Have Automobile Accidents*, by H. R. DeSilva. Published by John Wiley & Sons, Inc., 1942. P. 63.)

response in removing the foot from the accelerator, and the time elapsed until the brake was applied). The accuracy of steering the simulated car was also measured, and another reaction speed test was applied while the subject was engaged in steering ("vigilance reaction").

---

[18] H. R. DeSilva and T. W. Forbes, *Driver Testing Results*, Bureau for Street Traffic Research, Cambridge, Massachusetts, 1937.

It is difficult to present a representative sample of DeSilva's results with these tests. The differences and the degrees of significance varied from city to city in the various parts of his study. These variations are attributed by DeSilva to differences in driving conditions in the various cities, but we might also suspect some variation in the manner in which the test groups were selected. For example, the volunteer group which is used for comparison may have been a different kind of group in each city. Nevertheless the differences are all in the expected direction, even though some are not statistically significant.

More recently, Fletcher has used DeSilva's tests and some others in addition, in a comparison of "good" and "poor" drivers in California.[19] The good drivers included drivers of heavy and light trucks, and also a class of highway patrolmen, making a total of two hundred men in all. The poor drivers had been involved in a single accident resulting in a death, or in three or more accidents resulting in injury. Table 32 shows the whole series of test results upon these two groups of drivers. The first four tests listed would come under the heading of the present discussion—tests of reaction time and coordination. It will be seen that the differences, though small and indicating a considerable overlapping of the groups, are nevertheless statistically significant.

There may be some criticism of Fletcher's survey because of the specialized group of drivers chosen for the "good" driver class. Their test scores might not be considered representative of the performance of private car operators who are low in accident liability, and this might account for the more significant results which Fletcher obtained. Truck operators or others whose profession is driving might be a selected group with respect to the abilities tested, without the abilities necessarily being related to accident tendency. A deficient person might be able to compensate well enough for casual pleasure driving, but find driving fatiguing if he were to take it up as a full-time occupation.

Fletcher has considered the possibility that other factors have affected the test scores. He found that the groups differed in age and driving experience, as well as in test scores. The good drivers were younger and had driven a greater number of miles in the year preceding the test.[20] When these factors were controlled by selecting

[19] E. D. Fletcher, Preliminary report on special tests: Part I, Capacity of special tests to measure driving ability, State of California, Department of Motor Vehicles, pp. 1-12. (Mimeographed publication.)

[20] It should be noted that the mileage record of an individual has two implications. It implies greater practice in driving, but it also involves greater exposure

individuals of comparable age and mileage record, the differences in test performance still appeared. The poor drivers had the advantage of more years of driving experience, if that means anything for driving skill. This factor was also controlled in a special comparison, and was not found to affect the general results.

TABLE 32

FLETCHER'S COMPARISON OF TEST PERFORMANCE OF "GOOD" AND "POOR" DRIVERS *

Test		Average Scores "Good" Drivers	Average Scores "Poor" Drivers	Differences †	Critical Ratio ‡
Simple Braking Reaction (1/100")		37.73	39.31	1.58	4.2
Vigilance Braking (1/100")		47.26	53.44	6.18	9.0
Simple Steering (per cent score)		84.45	77.91	6.54	7.8
Vigilance Steering (per cent score)		76.42	71.42	5.00	4.6
Speed Estimation (error score)		2.20	2.47	0.27	2.7
Glare Sensitivity		3.5	24.0	20.5	7.5
Astigmatism (per cent failing)	Left	31.8	45.8	14.0	3.3
	Right	25.5	41.1	15.6	3.8
Acuity (per cent failing)	Both	1.0	5.0	4.0	2.9
	Left	7.0	21.8	14.8	5.1
	Right	7.5	15.3	7.8	2.9

* Fletcher, *op. cit.*, pp. 2-7.
† All differences are in favor of the good drivers.
‡ Critical ratio of 3.00 or greater indicates probability of less than .002 of obtaining the difference by chance.

Farmer and Chambers used a group of tests which they classed as "aesthetokinetic." [21] These included the "dotting" test described earlier in its application to the study of fatigue (page 81), a pur-

to risk. From the latter point of view, the poor drivers are more accident-prone than they at first appear to be, since their accidents, per mile, are still more divergent from the records of the non-accident drivers.

[21] E. Farmer and E. G. Chambers, A study of personal qualities in accident-proneness and proficiency, Industrial Health Research Board (Great Britain), 1929, Report No. 55. (With F. J. Kirk.) Tests for accident-proneness, *ibid.*, 1933, Report No. 68. A study of accident-proneness among motor drivers, *ibid.*, 1939, Report No. 84.

suitmeter test requiring the subject to follow irregular movements of a pendulum with a pointer, a two-hand co-ordination test in which the movement of a pointer is controlled by two handles to be manipulated simultaneously, and some other tests of similar nature. The tests have been used for workers in several skilled occupations, among

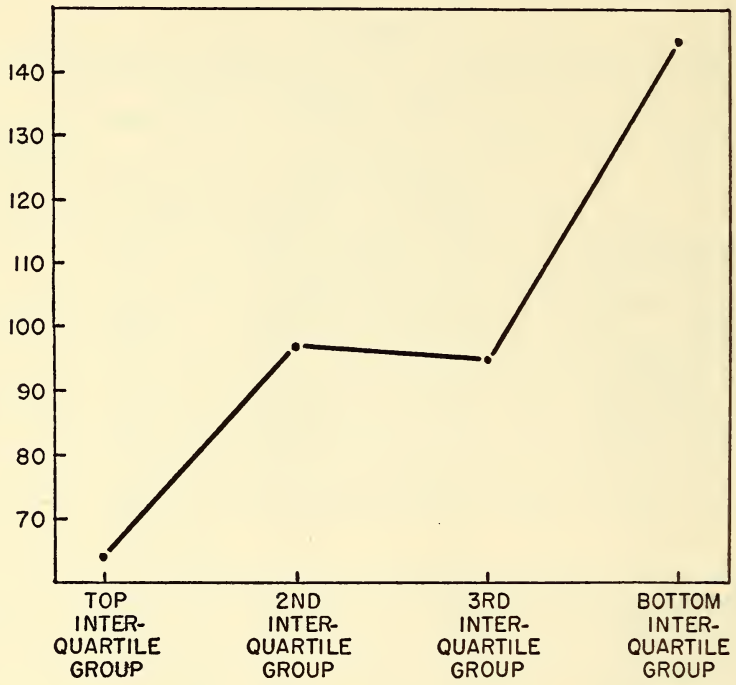


Figure 33. Accident Rates of Dockyard Apprentices, Related to Test Scores ("Aestheto-Kinetic Tests")

Difference between top and bottom groups is statistically significant.

(From Farmer, Chambers and Kirk, *op. cit.*, 1933, p. 21. Reproduced by permission of the Controller of His Britannic Majesty's Stationery Office.)

them shipwrights and carpenters, and also for bus operators. Figure 33 shows the accidents of dockyard apprentices who are grouped into quarters according to their scores on the aesthetokinetic tests. There were 251 subjects learning the trades of shipwright, electric fitter, and engine fitter. The top quarter of the group is significantly lower in its accident rate than the bottom quarter.

When Farmer and Chambers applied these tests to bus operators, the differences between the high-scoring and low-scoring drivers were not statistically significant. The investigators believe, however, that the tests may be useful in combination with other selective devices.

TABLE 33*

TABLE SHOWING THE MEAN PERCENTAGE ACCIDENT RATE OVER THE WHOLE PERIOD OF EXPOSURE OF GROUPS SELECTED BY DIFFERENT METHODS, TAKING THE ACCIDENT RATE OF THE WHOLE OBSERVED GROUP AS 100 PER CENT

	Shipwrights	Electric Fitters	Engine Fitters
Accident Rate of the Whole Group...	100	100	100
1. Accident rate of those left after rejecting the high-accident subjects in the first year	87	99	94
2. Accident rate of the top three inter-quartile groups of the aesthetokinetic tests	80	88	90
3. Combination of Method 1 and Method 2	75	88	79
4. Accident rate of the top inter-quartile group only in the aesthetokinetic tests	61	60	74

* Farmer and Chambers, *op. cit.*, No. 68, p. 28. Reproduced by permission of the Controller of His Britannic Majesty's Stationery Office.

TABLE 34†

EFFECT ON THE SUBSEQUENT ACCIDENT RATE OF REMOVING DRIVERS WITH MANY ACCIDENTS IN THEIR FIRST YEAR OR REMOVING DRIVERS WITH POOR AESTHETOKINETIC CO-ORDINATION

	Percentage of Drivers Removed	Percentage of Accident Rate	Percentage of Reduction
Accident rate of the whole group for all years except the first		100	
1. Accident rate after removing drivers with three or more accidents in their first year	28	91	9
2. Accident rate after removing the worst 25 per cent in the tests	23	93	7
3. Combining Method 1 and Method 2	44	87	13

† Farmer and Chambers, *op. cit.*, No. 84, p. 35. Reproduced by permission of the Controller of His Britannic Majesty's Stationery Office.

The possibility of selecting workers who will have a lower accident rate is suggested by Tables 33 and 34 from the same authors. The first of these tables illustrates a method of selection based upon the accident rate of an apprentice during his first year of work, along with the elimination of those with low scores on the tests. The other table, for bus operators, is similar, except that a test of perseveration

and a test of intelligence are included in the data. It will be seen that eliminating those drivers who have a bad record in their first year is slightly more effective than selection by tests.

2. Visual Tests. As the examples listed above would indicate, a great variety of visual characteristics have been tested in relation to accident susceptibility. It is probable that any visual defect is dangerous in certain occupations. On the other hand, the defective person can learn to avoid situations in which his defect will be troublesome or dangerous. It is therefore necessary to make direct investigations in order to determine the actual correlations between visual defects and accidents. DeSilva cites several sets of data indicating that poor visual acuity (inability to distinguish small details) is related to accident-proneness.[22] Weiss and Lauer, on the other hand, could discover no relationship,[23] although acuity did show some correlation to a laboratory driving test. Several studies have indicated that drivers who have one defective eye are likely to have intersection accidents more frequently than those with balanced vision. Table 35 shows the results of one such study.

TABLE 35 *

Vision and Intersection-Type Accidents

61	Accidents happened to defective-vision drivers on defective-vision side
10	Accidents happened to defective-vision drivers on normal-vision side
32	Accidents happened to normal-vision drivers
103	

* Fletcher, *op. cit.*

We shall not go into details about the other kinds of visual defect which may be studied. Table 32, page 273, illustrates some of the tests and differences discovered in the California driver study.

3. Distractibility and Carelessness. As a part of a battery of tests for selecting taxicab drivers, Snow designed a test of "emotional stability" which was found to be effective for selecting operators who would have good driving records.[24] The subject operated

[22] H. R. DeSilva, Why we have automobile accidents, *op. cit.*, pp. 71, 72.

[23] A. P. Weiss and A. R. Lauer, Psychological principles in automotive driving, *Ohio State Contributions to Psychology*, No. 11, 1931, p. 58.

[24] A. J. Snow, Tests for transportation pilots, *Journal of Applied Psychology*, 1926, *10*, 37-51; Tests for chauffeurs, *Industrial Psychology*, 1926, *1*, 30-45.

a kind of switchboard in which he was required to light a series of lights, in order, by manipulating a complicated arrangement of switches. He was also instructed as to what to do if a short circuit occurred. The short circuit, which produced sparks and a shock in the subject's hand, occurred a certain number of times in each test. With this arrangement it was possible to record the speed of reaction in dealing with the short circuit, and also to observe the changes in skin resistance which accompany certain portions of the test. The former measure of reaction speed was used alone for predicting performance of the drivers.

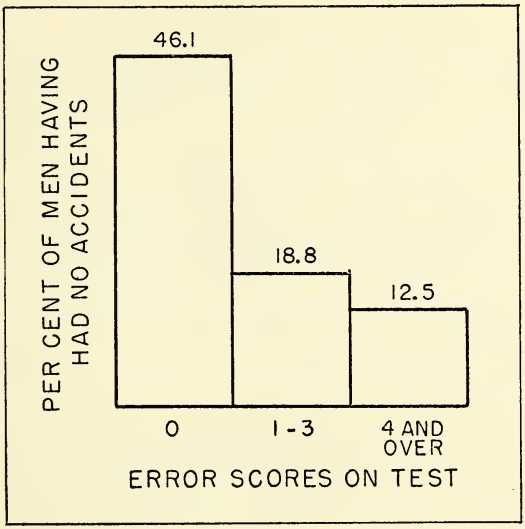


Figure 34. Relation of Accidents to Scores on Driver Tests (thirty-five cases) (From Wechsler, *op. cit.*, p. 27.)

Combined with an intelligence test, the test of "emotional stability" was effective in improving the accident records of operators hired for the Yellow Cab Company of Chicago. A group of men who were hired was subdivided into those who showed satisfactory performance on the tests, and those who were considered unsatisfactory. The satisfactory group averaged .2 accidents per man in a period of ten weeks, while the unsatisfactory group averaged 1.00 accidents during an identical period.

Unfortunately, we do not have comparable data for private car operators or for accident liability of industrial workers. Snow's success with commercial drivers would indicate the need for applying tests of this kind in studies of accidents in other fields.

In another investigation of taxicab drivers, Wechsler [25] employed a test involving a series of reactions performed with car controls in response to light signals. The test was designed to record not only the time of reaction, but also errors in reaction such as pressing the accelerator rather than the brake. Errors are considered a measure of "carelessness." The relation of these scores to accident rate for a group of thirty-five cases is shown in Figure 34. The reaction speeds showed a non-linear relation to accident record. As the speed increased up to a certain level the accident rate went down. If the reaction speed was especially high, however, the accident rate again increased.

4. Intelligence Tests. These tests have not been used as frequently as other tests in the analysis of accident-prone individuals. Some investigators report a correlation between intelligence and accident rate. Farmer and Chambers, however, could discover no correlation of any significance in their studies of dockyard apprentices.[26] The latter result may be due to a narrow range of intelligence among these apprentice groups, or it may be that intelligence is related to accidents only in certain occupations. Vernon suggests that we should expect accidents involving errors of judgment to be related to intelligence, but that the kind of accident suffered by the apprentices is more likely to involve skill rather than judgment.[27] When a correlation does appear, it may be spurious, however. The workers involved are in many different occupations, of differing risk. Consequently there may be a correlation of intelligence with the kind of work, and only indirectly with the accident rate. Only with hazard held constant can a correlation of a test with accident rate be meaningful. At present there is no good evidence that intelligence is related to accident liability in industry or in private car operation, although it may be an important measure for certain special groups.

Predicting Accident Rates by Means of Tests.—Most of the researches so far described have illustrated the possibility of obtaining significant differences in test scores between accident repeaters as a group and a group of accident-free individuals. Only the studies of Farmer and Chambers have furnished data upon the predictive value of the tests in sorting out individuals of high and low accident liability. Those studies, it will be remembered, indicated that the ac-

---

[25] D. Wechsler, Tests for taxicab drivers, *Journal of Personnel Research,* 1926, *5,* 23-30.

[26] Farmer, Chambers and Kirk, *op. cit.,* 1933, p. 30.

[27] H. M. Vernon, *op. cit.,* p. 44.

cident record of an individual for a year is of more value than the tests in predicting subsequent accidents. Other programs, like that of DeSilva, have used many more tests and more elaborate procedures. We should therefore consider the possible predictive value of these other programs, the components of which have been illustrated in the preceding pages.

Cobb, in a research carried out for the Highway Research Board, collaborated with DeSilva and with Lauer in an extensive analysis of the predictive value of the Harvard tests (DeSilva) and the Iowa tests (Lauer).[28] Some 3600 individuals took the tests, 2800 being given both complete batteries. In addition, information from the State Bureau of Motor Vehicles, and answers to a questionnaire or interview, were secured. Each item of information and each test was correlated with the accident history as given in the Motor Vehicle Bureau records.

TABLE 36*

Correlations with Accident Rate

Predictive Factor	
Miles Driven per Year	+.199
Education	−.190
Aptitude (intelligence)	−.195
Years Licensure	+.127
Drivometer Errors (a miniature test from the Iowa Battery)	+.112

* After Cobb, *op. cit.*

The highest correlations between individual items and accident history are shown in Table 36. Although there were thirty-three significant correlations obtained (for seventy-two tests in all), all those not mentioned in Table 36 were less than .10. That is, they were significant because of the large number of cases involved, but indicated little predictive value.[29]

[28] P. W. Cobb, Report to the Highway Research Board on the Automobile Driver Tests Administered to 3663 Persons in Connecticut, 1936-1937, and the Relation of the Test Scores to the Accidents Sustained. (Unpublished manuscript in microfilm filed with the Highway Research Board of the National Research Council, Washington, D. C., 1939.) Also summarized in H. M. Johnson, The detection and treatment of accident-prone drivers, *Psychological Bulletin,* 1946, *43,* 489-532.

[29] Cobb applies tests of significance which are strictly applicable only to correlations involving normal distributions, even though the accident distributions are known to be non-normal. We shall come back to this problem a little later.

There are two reservations to be kept in view in interpreting these individual correlation coefficients. First, a perfect test could not give a correlation of more than .454. This was computed by Cobb as the upper limit because of the unreliability of accident rate as an indicator of liability. Secondly, Cobb found that the state records of accidents were very incomplete, many of the operators having accidents outside of Connecticut which were not recorded in the Connecticut files. Thirdly, the group of individuals tested, though large, was also very heterogeneous, including a wide range of driving experience, of yearly mileages, of types of car operated, and so on. These are all factors which would tend to reduce the correlation between any test and the criterion of accident rate.

By means of a combination of predictive factors, Cobb was able to form a multiple predictive equation which gave a multiple correlation of .35, using all the tests (twenty-two) which would contribute anything to the prediction. He was also able to obtain a multiple correlation of .30 by a combination of factors which did not make use of any apparatus tests.

It is difficult to interpret these results, however, because Cobb has included among his predictive factors items like miles driven per year, years of licensure, and years of driving experience. The yearly mileage is a factor increasing the risk of accident, but the predictive value of tests would be most clearly demonstrated if the risk were held constant, allowing the individual differences in liability to appear in isolation from the external hazard. The other factors would be related to age, and also to possible changes in driving conditions over the years. Age and changing conditions should be kept as a problem distinct from that of determining the personal defects which contribute to accident-proneness.

Cobb also examined the following relationships: curvilinear correlations and critical scores on the tests; the possibility of "sifting" a group who were clearly deficient on several tests; and, finally, the correlations of specific tests with specific types of accident. These results were largely negative. Although a group could be "sifted" so that it consisted of those with a pattern of several deficiencies, this procedure showed less predictive value than the multiple equation described above. Almost no significant relationships between particular tests and specific types of accident were discovered. In the latter case, the failure may be due to the fact that Cobb examined only those possibilities which seemed reasonable *a priori,* for example the relationship of ability to judge distance to front-end and rear-end collisions. Perhaps the relationships to be sought cannot be worked

out *a priori*. This is made plausible by the fact that some of the relationships which he examined turned out to be the opposite of what had been expected.

Although Cobb's undertaking was on a large scale, it still leaves us with open questions. The results would, perhaps, have been more enlightening if he had used smaller, more carefully selected and controlled groups, rather than the large heterogeneous population studied. There is no essential contradiction between his results and those of DeSilva and others, since Cobb, like the others, obtained significant though small correlations. Cobb's results do indicate, however, that the further development and careful validation of driver tests is necessary.

The Uses of Tests for Accident-Proneness.—The preceding examples have given a few tests which are representative of a considerable number which have been used at one time or another. In illustrating the results of individual tests, we have shown that some of the tests have demonstrated value in selecting commercial vehicle operators and possibly for selecting workers in mechanical trades. Tests of this type have not been used for selection of private vehicle operators for several reasons. The most important reason is that the tests do not have sufficient validity for such widespread application. To prevent the owner of a car from operating it, the tests would have to have a very high predictive value before they could gain acceptance from the public at large. In commercial selection we may be willing to risk eliminating a few operators who might have been perfectly satisfactory because, in doing so, we are eliminating many whose liability is definitely high. The only criterion of success of the tests is the accident rate of those who are employed, and we can afford to run the risk of injustice to some of those who are refused employment. Obviously this procedure would not be acceptable for private vehicle operators.

Practical applications of tests like those described above to private vehicle operators, and in some instances to commercial employees as well, have been in the form of "clinical" treatment of the individual who has a poor accident record. That is, the tests are used as a method of diagnosing the reasons for high accident liability in accident repeaters. It is a process of individual treatment and recommendations based upon the particular individual's test scores.

DeSilva has reported the results of this kind of clinical diagnosis and advice in a group of accident repeaters who were followed for a period of nine months after they had passed through the "driver

clinic." Out of one group of 180 repeaters who came to the clinic in one city, 101 were tested and given advice, while the remaining 79 received some educational material of a safety campaign without the tests. Table 37 shows the accident records of these two groups before and after the time of coming to the clinic.[30] An even more marked effect is reported for a group of 470 repeaters.[31] The subsequent accidents for a period of six months were 90 per cent fewer in the tested group than they were in a similar group which was given "no tests and no education."

TABLE 37 *

RE-EDUCATIONAL EFFECT OF DRIVER TEST CLINICS

Comparison of accident record before and after clinic operation. Tested and untested repeaters. Nine months follow-up, Manchester, N. H., January, 1937

Number of Accidents Accruing to Groups of Repeaters Accidents per 100 Drivers	Untested (N = 79)	Tested (N = 101)
Before summons to clinic (15 months)	195	221
After summons to clinic (9 months)	38	18
Accidents after as per cent of accidents before 9 months in each case	32.5 per cent	13.7 per cent
Advantage of re-education by driver test clinic (compared to police attention, safety campaigns, etc., alone)		57.5 per cent

* H. R. DeSilva, The "why" of accidents, *Traffic Safety*, 1938 (July).

Johnson and Cobb have questioned the value of DeSilva's evidence as a demonstration that drivers' clinics are effective in reducing the accident rate in treated drivers.[32] Their criticism may be summarized in its application to the data of Table 37. The greater improvement in the tested drivers might be interpreted as (1) meaning that the two groups were different to begin with, and therefore might be expected to show different subsequent records; (2) that the clinical use of the tests is actually effective; or (3) that some other factor in the situa-

[30] No statement of the significance of this difference is given by DeSilva, and the present writer has been unable to discover a valid procedure for testing the significance because of the non-normality of accident distributions. Fletcher, whose similar studies are summarized on a subsequent page, reports a "critical ratio" for differences of accident rate, apparently basing his analysis upon standard statistical procedures which assume normality of the distribution of means.

[31] DeSilva and Forbes, *Driver Testing Results*, p. 49.

[32] H. M. Johnson and P. W. Cobb, The educational value of "drivers' clinics," *Psychological Bulletin*, 1938, *35*, 758-766. H. R. DeSilva, Automobile drivers can be improved, *ibid.*, 1939, *36*, 284-285. H. M. Johnson, Evidence for educational value in drivers' "clinics," *ibid.*, 1939, *36*, 674-675.

tion at the time of testing tended to reduce the accident rate.[33] With regard to the first of these possibilities, Johnson does not consider that DeSilva has shown the equivalence of the two groups before the time of testing. Whether this criticism refers to the difference in average accident rate between the two groups, or to other possible differences in variability within the groups, he does not state. Even if the differences were accepted as statistically meaningful, however, Johnson would still be in doubt as to whether (2) or (3) is the correct interpretation.

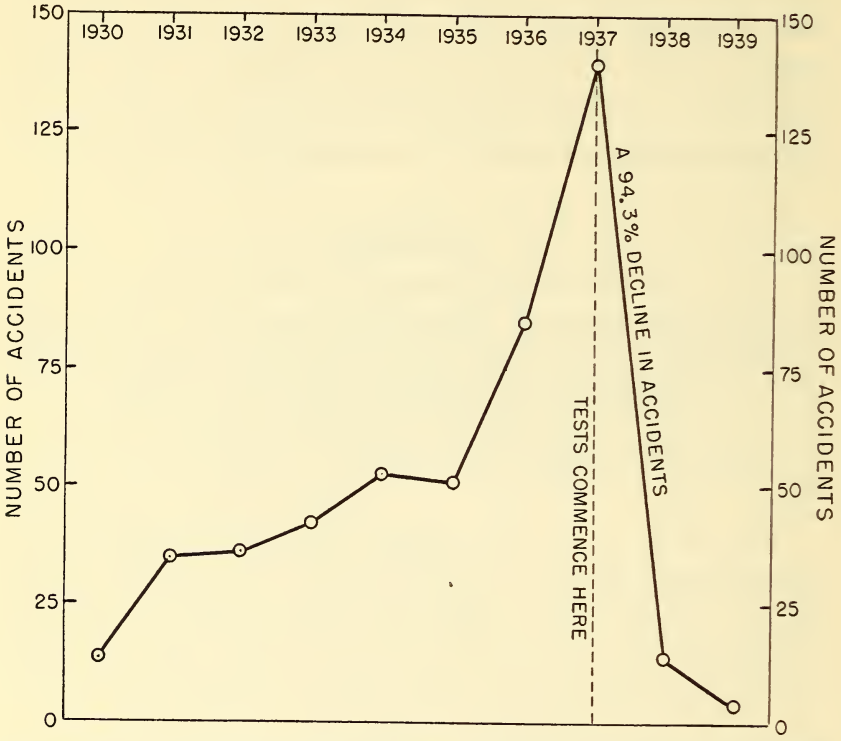
It does not seem that the differences in mean accident rate are sufficient to account for the different subsequent histories. According to DeSilva, there was no special selection of the drivers to be tested, and all had been sent to the clinic in the same manner. There is still the possibility that the change was not produced by the testing and advice which was given. Perhaps any tests, valid or invalid, would have the same result, somehow focussing the attention of the driver upon his driving habits. This is a distinct possibility, but again it does not seem that it would be sufficient to account for the whole results.

Fletcher has given more extended data upon the effect of the driver testing program.[34] The study includes several groups of truck operators, groups of state employees, accident repeaters whose accident rate had been increasing for a period of eight years prior to the test, and a number of drivers who had been warned because of repeated traffic violations such as speeding. In each case a similar group which was not tested was tabulated for comparison. If Johnson's criticisms of DeSilva are valid, they would probably apply also to Fletcher's studies. The comparison groups were not always strictly comparable in accident rate, and the *distribution* of accidents in each group is not given. In addition, there is not enough information upon his methods of determining the significance of changes to permit us to evaluate these procedures. Nevertheless, the changes resulting from the testing are so marked and consistent from group to group, that it scarcely seems possible to regard them as artifacts rather than bona fide effects of the testing program. In the case of the accident repeaters, for example, the accident rate declines within a year of

[33] Johnson and Cobb assume that it is possible to test for significance of difference of accident rates, but they fail to mention the fact that special methods are required because of the non-normality of the distributions involved. So far as the present writer has been able to discover, this statistical problem has not yet been adequately solved.

[34] E. D. Fletcher, Preliminary report on special tests (*op. cit.*). Part II. Effect of special tests on driving ability. (In mimeographed form.)

the tests to a level lower than that for any of the preceding eight years (see Figure 35).



Figures 35. Accident Reduction Following the Use of Special Tests (Accident record of 144 accident repeaters)

(From Fletcher, *op. cit.*, Part II, p. 12.)

The Generality of Accident-Proneness.—The topic of generality of accident-proneness has essentially the same meaning as the question of generality of any personal trait. It concerns the question of whether a person who is accident-prone in one occupation is also likely to be accident-prone in some other activity. Does the accident-prone driver also tend to have high accident liability in factory work? Does the person who is prone to minor mishaps such as cuts and scratches also tend toward greater liability to serious or fatal accidents?

The importance of answering such questions should be clear. If accident-proneness is a consistent trait in many circumstances, industrial employment departments will need to take each applicant's total accident history into account in selecting workers for hazardous occupations or even for occupations where the hazards are relatively

low. If proneness to different kinds of accident in the same work tends to be consistent in the same individual, safety engineers will wish to transfer workers who show trends toward minor injuries in order to avert more serious disaster. These are two distinct questions concerning the generality of the trait. They will be considered separately in examining the available data.

To give general answers to these questions would require a vast amount of research upon very large numbers of individuals over long periods of time. Consequently we can find only a few hints of the possible answers in the literature of the subject at the present time. We should have to raise these questions even if we had no information at all, however, since we must stress their importance as fundamental to the understanding of the whole topic of accident-proneness. Until more information of this kind is forthcoming, our handling of the question must be full of gaps.

Let us consider first the question of correlations between major and minor accidents within the same occupation or activity. Farmer and Chambers report correlations of this type based upon over fourteen thousand dock workers, including laborers, shipwrights, boilermakers, and the like, with the data covering a period of one year.[35] In various groups the correlation between minor accidents and major accidents (causing absence from work for a day or more) ranged from zero to .36. While the numbers involved are large enough to make these correlations significant in some groups, their small size may be due to the shortness of the period of observation. Here again we are using accident *rate* as an index of accident *liability*. The major accidents, being more rare, must be studied over longer periods of time before accident liability and rate are closely related.

Another more recent study also indicates a relationship between tendencies toward major and minor injuries. Jurgensen surveyed the data for 946 employees of one mill.[36] Among those who had disabling accidents, there was an average of 6.5 minor injuries in the year. Those who did not have disabling accidents averaged only 3.1 minor injuries. A similar relationship was evident when a serious accident was defined as an accident requiring the care of a physician. Among those who had such serious accidents, there was an average

[35] E. Farmer and E. G. Chambers, A psychological study of individual differences in accident rates, Industrial Health Research Board (Great Britain), 1926, Report No. 38.

[36] C. E. Jurgensen, How much do we know about accident causes? *National Safety News,* 1945, *52,* No. 5, 26-27, 80, 82, 84. Abstracted in *Psychological Abstracts,* 1946, *20,* 177, Abstract No. 1642. Original not seen.

of 5.5 minor injuries per person. Among those free of serious accident, there was an average of 3.0 minor injuries.

Until more complete data are available, it would appear to be best to assume that the two kinds of accident liability are closely correlated until we have clear proof to the contrary. While these statistical data leave the possibility of close relationships open, there is an important logical reason for expecting a correlation.[37] The same unsafe act may at one time lead to a minor scratch, at another the circumstances turn it into a serious injury. An individual is not liable to accidents as such, but has tendencies or habits which lead to the performance of certain unsafe acts. In some occupations, therefore, there may be a very close relationship between serious and minor mishaps because the same unsafe act can result in either.[38] In other occupations, however, the actions which lead to serious injury may be quite different from those which lead to cuts and scratches. While it is likely that the correlation will vary markedly from one occupation to another, it is safest to assume a close correlation rather than ignore the problem until the correlation is demonstrated. It should do no harm to institute transfers, retraining, or whatever else is necessary to prevent the minor injuries, and we may be preventing major accidents by doing so.

Another aspect of the interrelations of accidents concerns the correlations between accidents of different types in the same kind of work. For example, is there any tendency for a person who has a large number of falls to be specially liable to accidents involving machinery, or to cuts and bruises in the use of hand tools? In their study of bus operators, already referred to (page 275), Farmer and Chambers determined the intercorrelations between five types of driving acci-

---

[37] Tiffin (*op. cit.*, page 281) presents data to show that major and minor accidents are uncorrelated when the rates are determined by departments or occupations. This is, of course, a distinct problem. The factors *in a given job* which lead to minor accidents are not the same as those which lead to major injuries, as Tiffin properly concludes. These results have nothing to do with the *individual,* however, but with occupational differences in hazards.

[38] H. W. Heinrich (*Industrial Accident Prevention,* New York, McGraw-Hill Book Co., Inc., 1941, pp. 26-35) has urged that an accident be defined, not in terms of injury, but in terms of events which might produce injury. From various observations he has estimated that for each major injury resulting from such an event there are, in the long run, twenty-nine minor injuries and three hundred occurrences which do not result in injury at all. While these ratios will probably vary from occupation to occupation, the point is an important one. If records are kept in such manner as to include *all* falls, slips, and the like, their cause may be noticed and eliminated before the injuries occur. Heinrich was referring to locating causes of accident which would involve faulty tools, layout, lighting, or any responsible factor. The same reasoning may, however, be applied to the individual factors in accidents in many instances.

dents. The results of this analysis are given in Table 38. These authors conclude that some of these correlations are significant, and that there is a slight tendency for the different kinds of accident to be related to one another. Since the groups of drivers were not large to begin with, and the time period for Group B was only two years, it is quite possible that the small size of these correlations is due to random variations which would be eliminated with larger numbers and longer time-spans.

TABLE 38*

PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN DIFFERENT TYPES OF ACCIDENT

	Group A r	Group B r
Errors of judgment and overruns	0.367	−0.021
Errors of judgment and skids	0.148	0.049
Errors of judgment and miscellaneous	0.234	−0.001
Errors of judgment and blameless	0.161	0.045
Overruns and skids	0.077	0.082
Overruns and miscellaneous	0.233	0.097
Overruns and blameless	0.149	0.129
Skids and miscellaneous	0.161	0.074
Skids and blameless	0.214	0.060
Miscellaneous and blameless	0.390	0.075

* Farmer and Chambers, *op. cit.*, 1939, Report No. 84, p. 12. Reproduced by permission of the Controller of His Britannic Majesty's Stationery Office.

Herdan has suggested the application of factorial analysis to the interrelations of accident records.[39] For those familiar with Spearman's methods of factorial analysis, we may mention Herdan's results briefly. He first points out the similarity between determining the liability of an individual to accident, and determining the "G," or general factor, in intelligence tests.[40] When the Spearman method

[39] G. Herdan, The logical and analytical relationship between the theory of accidents and factor analysis, *Journal of the Royal Statistical Society,* 1943, *106,* 125-142.

[40] In fact, all the problems which we are discussing in this section could be treated by a variety of factorial methods, were sufficient data available. Spearman's methods would be satisfactory when they are applied to the problem of determining gross accident liabilities, with no need to differentiate between different kinds of accident. Where we are interested in determining whether or not there is a general factor underlying accidents of a number of different kinds, or in a number of different situations (which is the basic problem we are concerned with here), there is no one technique which can be said to be "correct." Thurstone's methods could be applied just as well, and group-factors rather than a single "G" factor might appear. Nevertheless it would appear that the field of accident statistics could profit by extended factorial studies using various techniques.

is applied, he finds that the miscellaneous accidents are most heavily "saturated" with the general factor of accident liability. Next, and equal, come the errors of judgment and the accidents in which there was no blame to the individual. Herdan considers this reasonable, since an error of judgment may complicate matters in a situation where the other driver is technically at fault. In other words, an individual may show his tendencies to accident in some accidents where the other driver is still at fault. The accidents involving skidding showed the lowest relationship to the general liability factor.

Herdan's results can, however, be considered as no more than an illustration of the possibilities of a method, since the data with which he works are unreliable, and the classification of accidents is probably not very satisfactory. In addition, other techniques of factorial analysis might have resulted in an entirely different picture of the results.

The relation of accident liabilities for the same individual in different occupations is still less known. One study has shown a correlation between the driving accidents of truck operators and their personal-injury accidents.[41] " 'Personal accidents' included not only those which occurred as the result of collisions in traffic, although these constituted an appreciable fraction of the total number. The larger proportion, however, were accidents which occurred in daily work not directly connected with driving; for example, many of these drivers were repairmen. Many of their accidents were caused by falling off their trucks; by dropping tools on their toes or other parts of the body; by mashing, burning, or cutting fingers; in general, accidents due to personal clumsiness." Table 39 shows the nature of this correlation of two types of accident very clearly.

Once more we find that our evidence is very incomplete. In this problem also there are logical reasons for expecting some degree of correlation between accidents of the same individual in different occupations. Certain defects of vision would have their effects in a number of different situations. The same might be said of almost any deficiency which is related to accident liability in any one field. A few deficiencies, however, might be specific to a single activity. Night blindness, for example, could not be expected to bear a relation to accident liability in most factories. Thus it is likely that a high degree of correlation of accident tendencies will be found in studying some pairs of occupations, while other pairs will show little relationship.

---

[41] U. S. Bureau of Public Roads, *The Accident-Prone Driver, op. cit.*, pp. 31, 32.

TABLE 39 *

PERSONAL ACCIDENTS OF DRIVERS FOR A PUBLIC UTILITY COMPANY COMPARED WITH MOTOR-VEHICLE ACCIDENTS OF THE SAME DRIVERS: 1930-1935

(Average personal accidents per operator grouped in accordance with the number of motor-vehicle accidents each in the same period.)

Number of Motor-vehicle Accidents per Operator 1930-1935	Number of Operators	Personal Accidents per Operator	Relative Risk
0	81	0.765	1.00
1	44	1.273	1.66
2	68	1.721	2.25
3	41	2.222	2.91
4	25	2.200	2.87
5 and more	54	2.556	3.34
Total	313		

* U. S. Bureau of Public Roads, *op. cit.*, p. 32.

Again, however, it would be best to assume that there is correlation until it has been proved otherwise. An individual with a record of a high rate of accidents in any field should not be placed in a hazardous occupation until more information is available. The further information which we would need would be either the demonstration that there is no correlation of occupational accidents in the fields involved in this particular case, or test results which show that the individual's deficiencies will not prove a handicap in the occupation for which he is being considered.

"Personality" and Accident-Proneness.—It is frequently suggested that certain emotional or temperamental traits contribute to accident liability in certain individuals. In fact it is sometimes implied that *most* accident-prone individuals are high in liability because of these general and ill-defined "personality" traits, rather than because of specific deficiencies in sensory or motor equipment. If this is true, we certainly should expect a high degree of generality of the trait of accident-proneness, since it would depend very little upon the specific activities and performances of any given job. It would also mean that we have a long way to go in developing means of diagnosing accident-proneness.

We are not in position to say how much of a part these "personality" characteristics play in accident liability, since we do not have as yet any clear statement of what personality traits are involved,

nor do we have any very satisfactory personality tests to work with.

There are suggestions of the importance of temperamental factors in some of the available accident statistics. Vernon found, for example, that accident rates of individuals correlated with their visits to the first-aid room for treatment of minor ailments, although there was no correlation between rate of serious illness and accidents. Since minor ailments are frequently correlated with neurotic tendencies, the latter may be the responsible factor in determining the correlation. Still more indefinite clues appear in the clinical studies of the accident-prone in which there is always a group which is classified under the heading "faulty attitude," the investigators having failed to discover any more definite causes of accident tendency.

In recent years Dunbar and her collaborators in the field of psychosomatic medicine have been collecting psychiatric case studies of hospital patients under treatment for fractures.[42] Originally these patients have been chosen as a "normal" control group for studies of other groups of patients, *e.g.,* those suffering from heart disorders. The investigators found, however, that many of these patients were accident-prone, and that they displayed a high frequency of previous accidents. They also came to the conclusion that there was a composite personality pattern which could be derived from analysis of their case studies of fracture patients, a pattern distinct from that found in other groups of patients.

Before summarizing the details of personality characteristics of the accident-prone group, it would be well to examine the point of view of the investigators as they approach their data. Their conclusions must be based upon voluminous notes of interviews and personal histories. There is, therefore, a definite chance of the investigator's becoming set to look for certain elements and thereby oversimplifying or distorting the results.

These investigators assume first of all that there is proof that tests based upon co-ordination, reaction time, driving skill, and the like are invalid. The data which we have presented would justify no such assumption. It is true that the validities so far observed have been low—in some studies much lower than in others. Nevertheless we have already given reasons why these validities would be low even if all accident-proneness were due to specific deficiencies like

---

[42] H. F. Dunbar, T. P. Wolfe, E. S. Tauber, and A. L. Brush, The psychic component of the disease process (including convalescence) in cardiac, diabetic, and fracture patients: Part II, *American Journal of Psychiatry,* 1939, *95,* 1319-1342. F. Dunbar, Medical aspects of accidents and mistakes in the industrial army and in the armed forces, *War Medicine,* 1943, *4,* 161-175. H. F. Dunbar, *Psychosomatic Diagnosis,* New York, Harper & Bros., 1943.

poor dexterity or reaction time. Poor validity of these tests does not prove that the factors causing accident-proneness must be sought in general personality characteristics. Dunbar appears to have exaggerated the poorness of the relationships in previous studies by concentrating her attention upon the negative findings. This assumption does not necessarily lead to outright errors, since all would admit that general personality factors will be important in a certain proportion of accident cases. It may lead, however, to a mistaken emphasis upon these temperamental characteristics at the expense of careful study of other possible factors.

We should have to know more details of the methods used by Dunbar and her co-workers if we were to weigh the possibility that a "type" has been imposed upon the data by selection. Although Dunbar speaks of certain subtypes of accident-prone individuals (based upon differences in diagnosis by the Rorschach Test), the picture which is presented is that of certain common features running through most of the group, with the subtypes representing variations in certain specific traits. We should have to see more definite quantification and statistical treatment of the distributions of these traits before we could determine the consistency of the patterns. As a simple example, suppose it is found that the fracture patients show two traits in unusual frequency; very strict parents and tendency to truancy from school in childhood. Does this mean that there is one group in which the first trait is very frequent and is related to accident frequency, while there is another distinct group of individuals who display the second trait in high frequency? Or does the result mean that the fracture patients tend to show a pattern in which the two traits appear *in combination* in abnormally high numbers? The investigators applied statistical treatment of individual traits, but the discovery of patterns of traits appears to have been carried out on the basis of general impression, coupled with certain theories of personality organization derived from psychiatric practice. Yet the patterning is as important as the discovery of single traits, since entirely different methods of diagnosis and prediction would have to be applied in the two cases.

With these cautions in mind, we may attempt a brief summary of the personality traits which Dunbar found to be correlated with being a member of the fracture group. Since most of the comparisons involve the prevalence of the trait in the fracture group as compared with the group suffering from coronary occlusion, some of the "traits" may not be special characteristics of the fracture group. Instead, they may frequently represent traits in which the coronary

group is abnormal, while the accident group is normal. In some instances it is possible to refer to a general norm; in others, no general norm is available.

The following are some of the more clear-cut distinctions between the two groups, briefly noted:

1. Higher accident history in family and siblings of accident cases.
2. Better health record.
3. Tendency to interrupt education, although intellectual level is normal.
4. Some accident-prone tend to stick to jobs, but without attempting to advance; another group shows job mobility. Coronary group sticks to one job and tends to "work to the top."
5. Higher proportion of early neurotic traits, such as lying, stealing, truancy, sleepwalking. Later, almost no obvious neurotic traits. Reverse tendency in coronary patients.
6. Tendency to take stimulants to "let off steam," while coronary group may take them to keep on working.
7. Greater interest in sports among fracture patients.

Beyond these relatively specific comparisons, Dunbar also presents a more qualitative differentiation of temperamental characteristics. She depicts the accident-prone individual as being impulsive—the tendency to *do something* when involved in an emotional situation. The cardiac patients did not show this tendency. The accident patients try to escape from authority, including that of strict parents, while the cardiac individuals try to overcome or subdue the authority.

Dunbar has proposed the use of selective procedures in the armed forces based upon the patterns we have just sketched. The information would be gained from the history of the individual, from questioning him, and from observing his behavior during induction or training.[43] Thus far, however, the predictive value of such a procedure is still to be determined. It is one thing to find distinguishable patterns of traits among those already known to be accident-prone, it is quite another thing to pick out the accident-prone from a heterogeneous group of inductees or job applicants.

The work of Dunbar is the only study of this problem of a systematic kind.[44] In the literature on psychoneurosis there are

---

[43] *Op. cit., War Medicine.*

[44] Johnson (*op. cit., Psychological Bulletin,* 1946, 518-524) refers to the psychiatric approach to accident problems, pointing out the lack of even elementary statistical analyses in instances of this approach which he has observed. He does not, however, take account of Dunbar's work in his discussion.

occasional references to individuals who had accidents because of some particular neurotic symptoms. Culpin describes several such cases in a discussion of accident-proneness, but the cases came to his notice incidentally in general psychiatric practice.[45] Hersey also gives instances of "emotional" causes of accidents and accident-proneness.[46] Dunbar's analysis would imply that the ordinary kinds of recognized neuroses are not an important factor in the adult accident repeater. If this is correct, devices for diagnosing neurotic tendency would not be effective in weeding out the accident-prone from dangerous occupations. Nevertheless it would be worth while to try out these diagnostic methods when and if measures of satisfactory reliability and validity are available. Certain kinds of neurotic pattern may entail accident tendencies as a part of the expression of the neurosis. Others may be accident-prone for temperamental reasons, but not classifiable as neurotics in the customary meaning of the term, like the cases described by Dunbar.

Conclusions

Needs for Research in Accident-Proneness.—As in most fields of research, when our knowledge of a field increases, we become more and more aware of the important gaps in our information, and of the need for new types of research which could not be envisaged before our present knowledge was obtained. Some of the information which we need can be gained from further statistical analysis of accident data itself. Other problems require the development of more adequate procedures of analysis of individual differences, and further study of their correlations with accidents.

The information which we need to derive from the accident statistics consists largely of correlations. As we have seen, we wish to know more of the generality of the accident trait. Does the individual tend to be prone to only one kind of accident, or is he prone to accidents in general? Does he tend to have accidents only in a certain kind of situation, or do they occur in a great variety of circumstances? Is the seriousness of the outcome of an accident a matter of chance, or does accident-proneness express itself in leanings toward minor accidents in abnormally high or low proportion? That is, does accident liability express itself in tendency toward a

---

[45] See the discussion of Newbold's paper in the *Journal of the Royal Statistical Society,* 1927, *90,* 540-544.

[46] R. B. Hersey, Emotional factors in accidents, *Personnel Journal,* 1936, *15,* 59-65.

certain frequency of accident without regard to the seriousness of the result, or can a person have a high liability to minor accident coupled with a low liability to serious accident? As we have seen, we already have some hints on these problems, but there is a field for much profitable research in extending our knowledge of these problems.

With regard to the individual characteristics which may be correlated with accident tendencies, we have already seen that the procedures so far in use have been relatively crude—correlating test scores against over-all accident frequency. Some of the difficulties we have found in these procedures might be avoided by more elaborate methods of analysis. For example, Fletcher's report upon an analysis of intersection accidents, mentioned on page 276, suggests a pattern of analysis which would make test data much more meaningful. Instead of comparing accident repeaters *en masse* with accident-free groups, it would appear to be much more profitable to subdivide the accident repeaters into various subgroups and work out the relationships separately for each group. For example, the first step might be a study of those involved more than once in a single type of accident (intersection, collision in passing, running off the road, and such).[47] These divisions should be made as specific as possible. Then, if several types of accident appear to be correlated with the same pattern of test scores, these accident types could be grouped together.

Another approach would consider the problem of *patterning of deficiencies* as a factor in accidents. It is conceivable, for example, that a person who is somewhat deficient in *two* traits, such as reaction time and speed estimation, is much worse as a driver than a person

[47] It may be suggested that Cobb has already tried this kind of analysis without success. (See page 280 above.) We have seen, however, that his analysis was incomplete. He tested only for correlations which appeared "reasonable" on a common-sense basis. There may be other relationships which could not be guessed in this manner. The proposal made here is that the results of many different tests be examined for each type of accident. Certain individual tests or certain combinations of tests might have a predictive value which is greater for the specific kind of accident than it is for accidents in general.

Another difference between this proposal and Cobb's method is that Cobb correlated scores on a given test with frequencies of accidents of a certain kind. We suggest here a selection of individuals who seem to be consistent in the kind of accident which they suffer, and a comparison of their test scores with those of an unselected group. Just as accident rate is partly a function of chance, so the type of accident may not always reflect the personal tendencies of the individuals involved. If an individual has two or more accidents of a similar kind, however, a personal factor determining the kind of accident is more likely. Such an analysis would be, in effect, a search for specific factors in accident tendencies, as contrasted with the problem of more generalized accident-proneness.

who has a single severe deficiency. Are there any patterns of deficiency which are related to very high accident liabilities? The analyses now available do not answer such a question. In the Harvard clinics, for example, each test was analyzed separately. Farmer and Chambers used combinations of tests in their predictions of accident rate (see page 275) but this was merely an additive combination. There is no way of determining from such an addition of scores whether a particular combination of deficiencies has an effect which is more than the sum of the deficiencies.

Dunbar's suggestions about personality factors will need to be followed up with further validation and more concrete procedures for diagnosis if we are to take the "personality" factors into account in predicting accident tendencies. As she herself has suggested, there is no adequate control group of low-liability individuals which can be compared with the accident-prone with respect to the psychiatric patterns she describes.[48] The ability of a psychiatrist to pick out of an unselected group those who have an accident tendency has not as yet been demonstrated. The therapeutic value of treating these cases has not been demonstrated either; in fact, there is less evidence for improvements from this form of treatment than there is for improvement as the result of specific aptitude testing.

This brings us to a final need. This is the more thorough testing, in a statistically unimpeachable manner, of the therapeutic value of driver tests, and the trying out of similar clinical techniques for accident repeaters in industry.

All these suggestions call for years of research by a variety of specialists. It is hoped that the interest of governmental agencies and of industry in these problems will continue and develop, and that those who are working in the field will not overlook new approaches in the form of new testing methods which have not heretofore been developed.

Suggestions for Practice.—The knowledge of the problem of accident-proneness which we now have can contribute to practical approaches to accident prevention, even if there are many loopholes in our information. The safety department of a plant, the personnel department concerned with transfers of workers or hiring new employees, the state highway authorities, can not only support research in the future but they can also make use of what has been done.

---

[48] H. F. Dunbar, *Psychosomatic Diagnosis,* New York, Harper & Bros., 1943, p. 690.

The following procedures are recommended as in keeping with our present knowledge of this field:

1. Record keeping for accident reduction.
   (*a*) Detecting the accident-prone. As we have suggested, it is safest to assume generality of the accident trait. If we assume that a person with a high rate of minor injury is more likely to have serious accidents, such a person should either be transferred to a less dangerous position or given remedial treatment if that is available. The personnel department should make every attempt to find out the accident history, major and minor, of a prospective employee, including off-the-job as well as industrial accidents.
   (*b*) Detecting external factors in accidents. Accident records should be kept not only in terms of the individual involved, as we have just suggested, but also in terms of the location, the cause, and all other pertinent surrounding conditions. As Heinrich suggests, this may also prevent serious accidents later by showing up factors in the working situation which make for "unsafe acts."

2. Use of all available testing methods for hiring men to work in occupations with a high injury rate, and for diagnosing those who are already employed and who show an accident tendency.

This would provide further information for validating the tests, and in addition would have a beneficial effect upon the worker taking the test. The tests might have this latter effect even if they have low validity—the worker may give more attention to possible habits or shortcomings which are making him accident-prone.

Chapter 13

SKILL AND PRACTICE

In a general survey of the topic of efficiency, learning is both a factor contributing to efficiency of performance and an activity which itself provides problems of economy in time and effort. There is little doubt that practice results in marked improvements in efficiency of any given performance, provided that the worker is being taught effective methods of doing the job. In muscular work the metabolic cost per unit of output is lower in the skilled athlete or laborer than it is in the novice. Where indices of effort in sedentary work have been available, similar results have been obtained. In fact it is scarcely conceivable that practice could reduce efficiency unless the training is so ill-conceived that the worker is learning methods which are awkward and wasteful. Even when the method of work which is being taught is not the best method which could be developed, the practiced worker carries through the procedure with fewer waste motions, fewer interruptions and errors, and he feels less strained. Again, accidents are usually very frequent during the novice stage of practice.

Practice and training themselves require effort and take up the time of the worker and his instructor. Since improvement in performance involves a cost to the trainee, we are again faced with problems of efficiency. We wish to determine, therefore, the method of practice and the method of instruction which produce the maximum improvement for a given amount of effort and time.

In most writing upon the subject of efficiency of learning, the only element of cost which is taken into account is the *time* consumed in learning. In dealing with work in general we have found it a serious misuse of the term to confuse efficiency with rate of performance. While it is still misleading to confuse rate and efficiency of learning, there is probably more justification for the interchange of terms in connection with learning than there is in other aspects of work. For one thing, the trainee is using not only his own time and effort, but usually the time and effort of an instructor also. If the trainee adopts a method of learning which is slow but easy for him, it may

require more effort by the instructor. In any case, the time factor becomes doubly important. Secondly, the main danger in overlooking the total cost of work in the ordinary working situation is that the additional cost may lead to an accumulated deficit, as also to a long-term deterioration of the worker, with ultimate loss of productivity. During the training period, however, it may require more effort for the trainee to learn quickly, but this is only a temporary condition which is not likely to lead to harm in most jobs. Then, too, the trainee arrives at a stage of practice where the task itself, as distinguished from the practice, requires less and less effort. In other words, inefficiency due to high effort level in early training is a temporary inefficiency which may be more than balanced by the later efficiency of the worker.

For the above reasons we do not need to be seriously concerned about the confusion of rate and efficiency of learning. The confusion does exist, however, and it would be clearer and more honest to speak of the *rate* of learning in discussing most of the research that is available in this field. We certainly know very little about the true efficiency of learning or the factors which affect it.

The amount of published material upon the subject of methods of practice and training is enormous. It consists of information falling into several categories: (1) Laboratory studies of single factors in isolation. (2) Factory experiments in which single variables are studied but others are only partially controlled. (3) Experiments in the schools dealing with the learning of academic subjects. (4) Discussions of "practical" training programs in which the principal emphasis is upon the administration of the program, *e.g.,* such topics as how the instructors are chosen, the content of training courses, and the like.

Although each kind of information has its special value, the various kinds are by no means equal in usefulness. The fourth group of "practical" discussions appeals to industrial managers and supervisors, but frequently it tells us almost nothing about how the rate of learning can be improved. Often there are no quantitative data to show the effect of the training program. When such data are given, the program is so complex that we cannot tell what factors were responsible for the improvements which occurred. Often the procedures which are supposed to have increased the speed of learning are special tricks which would apply only to that special job. They are therefore of importance only to those who are training workers for the identical occupation.

Of the other data that are available, the scholastic experiments frequently involve a situation in which the motivation is so different from that common in industrial training that it makes it extremely difficult to transfer the results. The same criticism also may be applied to a lesser extent to the laboratory experiments. In any case, these methods provide at most a suggestion which must be checked later by direct factory experimentation.

The ideal approach to the development of an industrial training program would consist of drawing up recommendations from the material already available in the field of learning, and then testing those recommendations by means of an experimentally controlled trial. In more detail, the procedure might be outlined as follows:

1. Determination of the best method of performing the work; that is, finding out in detail what is to be learned. This problem was considered in Chapter 7.
2. Determination of the sequence in which the aspects of the job are to be taught, ignoring the traditional method of teaching, and considering experimental information which might have a bearing upon the ease of learning the material in various orders and in various methods of subdividing the job.
3. Drawing up a detailed teaching and practice plan. This plan would contain:
   (*a*) Descriptions of the instructor's technique in introducing the worker to each phase of the job.
   (*b*) Scheduling the practice sessions, including their duration and spacing, and the type of practice which is to be carried on during each session if possible.
4. Explicit provisions for motivating and developing interest. (*E.g.*, instructional material which would make the job meaningful even though the material is not essential to the operation in itself, provision for incentive pay during training, and the like.)
5. Selection of a suitable measure of proficiency. (If the performance itself is not reliably measurable, it will be necessary to use other forms of achievement test. See a general textbook on industrial psychology for a discussion of achievement testing.)
6. Provision of a control group to be trained by former methods, or by other methods under consideration. The control group of workers should be as nearly identical with the experimental group as possible in each of the following qualifications:
   (*a*) age,
   (*b*) aptitude for the work,

we mean by recollection, the latter is simply a result of rote memorizing. Although recollection is not an explicit aim of most forms of industrial training, it is an aid which the learner frequently uses —he recreates the diagrams of his textbook, he listens again to the instructions of his teacher, and so on. It is a component of learning which we cannot now control because it is so little understood. It appears to follow completely independent rules, and to be dependent upon many factors other than those which control the four types of learning we have listed in Table 40. It must be kept in view because it may account for some of the apparent irregularities which occur in learning. Individual differences in learning under a given set of conditions are often due to the differences in individual ability to recollect in this manner. Also, two tasks which appear to be very similar in nature may not respond to the same conditions of learning because of the different involvement of recollection in the two situations.

TABLE 40

KINDS OF LEARNING

1. Verbal memorizing.
2. Learning of a fixed motor pattern.
   (For many purposes this class may be grouped with the verbal type of memorizing. Examples would be typing and machine operation of some kinds.)
3. Learning *dependent* upon comprehension.
   (The task involves mechanical principles, understanding of an accounting system, or the like.)
4. Learning an adjustive or variable motor pattern.
   (Operating a car or airplane, learning a sport like tennis. What is learned is not a fixed sequence of movements but a pattern which constantly varies with a changing situation.)

As we proceed to consider the things which the training supervisor must keep in mind as he plans, we shall discuss only those factors which have been investigated experimentally. We shall have to consider not only the presence of experimental evidence in support of a given recommendation, but also the extensiveness of this experimentation in the various classes of learning we have described. A factor which has been found to be beneficial in a number of experiments upon rote memorizing may legitimately be recommended for application to other tasks involving rote memorizing, subject to the experimental checks which we have already described. To extend the

recommendations to tasks involving other classes of learning is a more doubtful procedure. It is worth the trial in many cases because such a trial extends our knowledge of the subject of learning. In terms of the immediate practical aims, however, such a recommendation is not to be considered much more valuable than ordinary common-sense hunches and guesswork.

On preceding pages we have outlined the task of the person who must plan a training procedure. In presenting the evidence upon factors in learning, we shall follow this same outline. It would be easier to survey what is known without reference to the details of the planner's task, and it would give the impression that a substantial amount of information is available. Even though there is a background of hundreds of experiments upon learning, there are many gaps in our knowledge which appear clearly only when we try to match our experimental results with the task of the training supervisor.

Our approach may not do justice to the body of scientific information upon learning now available, but it will show the applicability more clearly, and it will also bring out the need for much further research.

Factors in Determining the Subdivision and Sequence of Material.—It might appear that most subjects to be taught, whether in a factory or in a school, have natural subdivisions and natural orders in which the various subdivisions may be taught. It is true that many of these divisions and sequences are fixed by the nature of the task to be learned. Many so-called "natural" divisions, however, may be only the result of habit and tradition on the part of the teacher. Under these conditions, the planner of a teaching program should not be misled by the apparent naturalness of a sequence into ignoring other non-traditional possibilities which may be more effective than those now in use. For example, it might be that the traditional method of teaching a person to drive an automobile by spending a long time in practicing the manipulations of clutch and shift lever at the beginning of practice is not the most effective one. Perhaps the person would learn more rapidly if he were to spend very little time on these manipulations at first, and were to spend more time practicing steering and driving on the open road before perfecting his skill in shifting. We do not know which would be the more effective until both have been adequately tried out. Belief or disbelief upon the part of practiced teachers in the effectiveness of a method is insufficient evidence to settle the question.

Many problems of this sort can be settled only by research upon the job in question, following the outline we have suggested. Certain general features of subdividing the learning material have, however, been investigated in a sufficient variety of circumstances to permit some generalization. These generalizations should provide at least some background for operation by the person charged with laying out a training program.

Size of the Practice Unit. One general factor which has been investigated quite widely is usually labeled as the "whole-part problem." In almost any task of learning, there arises the question as to how much of the material should be mastered before the trainee passes on to the next segment. In learning a language, should the individual master thoroughly a very few words, then a few more, and so on, or should he be given a large number of words all at once? If a job requires co-ordinated manipulations with both hands, as in playing the piano, should the trainee master the technique with each hand separately before attempting the combined operation?

This question has been investigated with verbal lists, poems or numbers to be memorized, and also for certain muscular skills. In most cases there is an advantage for the "whole" method, that is, for exposing the trainee to a large amount of information, or to a more complex co-ordination, rather than subdividing material into parts which are (individually) more easily mastered.[1] There are limits, of course, to the amount which can be placed in a learning unit, but the optimal amount of material, as determined in experimental studies, has been larger than that which would naturally be chosen by the trainee or by most teachers. It should also be remembered that the "whole" method can often be combined with special practice upon "parts," with results which are superior to those obtained with the complete emphasis upon the whole task.

Although the principle of "whole" practice is quite general in the sense that it has been found effective in a variety of learning tasks, there are certain conditions in which the "whole" method becomes less effective. In verbatim memorizing it appears that the "whole" method is itself a skill which must be practiced. Subjects in some experiments showed an advantage for "part" practice early in the experiment, but later the "whole" method gave better results. There is

---

[1] For a survey of experimental results see R. S. Woodworth, *Experimental Psychology,* New York, Henry Holt & Co., 1938 or J. A. McGeoch, *The Psychology of Human Learning,* New York, Longmans, Green & Co., Inc., 1942.

also some indication that children profit less by "whole" practice. This point may be especially important in certain kinds of industrial training where the trainees are unaccustomed to systematic learning.

Most of the experimentation upon the factor of "whole" learning involved tasks which belong to the first two types of learning mentioned previously (page 302), and some refers to the fourth kind. Little or no experimental data are available about the effect of such a factor in learning which is dependent upon understanding. It is often assumed that the "whole" method should be applied to this kind of learning as well; for example, in scholastic work. Such an assumption is not justified, since there is no reason to suppose that this kind of learning depends upon the same factors as other kinds. Probably we shall ultimately find that the "whole-part problem" is much more complex when we turn to comprehending.

Nevertheless, we shall probably find that some modification of the "whole" method is a good general rule even in learning that is dependent upon understanding. Each segment of a field is more meaningful and more readily learned if the student has at least a general notion of the place of that segment in the total field, and some idea of the interrelations between the thing learned and the other elements in the same total framework.

We must emphasize that the preceding statements are speculative, and this is not supposed to be a speculative book. The speculations are inserted, however, in order to suggest the need for more adequate research upon the conditions of learning which is dependent upon understanding, certainly a very important form of learning not only in the school but also in industry.

Factors in the Sequence of Training. Variations in the possible sequence of learning different aspects or segments of a job are potential factors in the rate of learning, but they have been given only isolated consideration. In rote memorizing or in learning a fixed movement pattern, the sequence itself is the important thing to be learned, and usually cannot be varied for that reason. In adaptive learning and in learning dependent upon understanding, and even in the more mechanical forms of learning where the task consists of several distinct phases, sequence is a potential variable. Unfortunately, these are exactly the kinds of learning which have been most difficult to study experimentally.

One problem which arises where there is a choice of sequences is, however, amenable to experimental study. This is the problem of

transfer and *interference.* If, for example, it is decided to train the worker in the simpler phases of the job first, we must consider whether this initial training is likely to be an advantage or a disadvantage when the trainee comes to the later phases of his training. We may sometimes wish to begin the training with a simplified version of the job, on the assumption that this will make later practice in more complex operations much easier to master. Similar considerations arise when we consider the transfer of workers from one task to another.

The positive transfer of skill from one manual task to another, or from one sensory discrimination to another, is much more limited than is believed by many laymen. The mere fact that one job requires dexterity in finger manipulation (for example) does not guarantee that the worker will learn a new task involving the same general type of bodily movements.

Two major conditions for positive transfer have come out of experimental studies. First, transfer occurs if a large proportion of the new task is made up of movement patterns which are *identical* with those in the task learned earlier. We shall presently give an illustration of the fact that mere general resemblance of the movements is not sufficient, and that identity must be interpreted in a very narrow or rigid sense. The second condition favoring transfer is the existence of common principles underlying the two tasks—if these principles are explicitly taught or discovered by the subjects. In other words, the trainee must understand that the principles apply to both tasks if they are to lead to transfer.

An experiment by Cox illustrates these two points.[2] The two tasks both involved assembling and disassembling parts of an electric fixture. The only difference was that the two tasks involved different parts of the fixture. Yet the experimental group which practiced first one task and then the second did not learn the second task any more rapidly than the control group which practiced only the second task. This was the result for bare "practice" as contrasted with what Cox called "training." In the latter case, the experimental group not only practiced the first task but also was given explicit instruction in methods, such as recommendations upon the arrangement of materials and the distribution of attention. With this additional factor of instruction in general principles of workmanship, a reliable transfer effect was obtained.

---

[2] J. W. Cox, Some experiments on formal training in the acquisition of skill, *British Journal of Psychology,* 1933, *24,* 67-87.

Thus we should be wary of attempts to prepare the trainee for complex manipulations by training him first in simpler manipulations, unless the simpler task is *identical* in movement pattern to portions of the more difficult job, or unless the simpler activity is used to teach *explicitly* important general principles involved in both. If the simpler task is introduced solely as a training device, and if it is not a part of the job as carried on in practice, we may seriously question whether the introductory training saves anything in the long run. In some instances the gain may be illusory, due to the apparent ease of beginning the complex task after this introduction, but the overall time may be no greater if the trainee starts immediately upon the complex task without any introductory practice. The planner of a training program must seek to avoid such illusory gains and to search for the procedure which will give most adequate results *in the long run.*

It may be argued on common-sense grounds that these illusory gains are important in building up the self-confidence of the worker and keeping him interested in the learning. Such a claim can be substantiated, however, only by direct trial. It can be just as logically argued that initial difficulties in learning would have a beneficial effect upon motivation. Engineering students, for example, are proud of the fact that their work is difficult and that many cannot learn it. A method of learning which sacrifices speed of learning in order to make the progress of the trainee deceptively smooth might be less challenging and therefore less interesting than a method which introduces complexities early. One of the experimental problems awaiting the further study of learning is involved here, and the optimum condition is still unknown.

In other phases of planning the sequence, the possibility of *negative* as well as positive transfer must be kept in view. In general, the greatest negative transfer or interference occurs where the two things to be learned are closely similar in their general nature, yet contain some disparate elements. When tasks of this nature are practiced one after the other, the interference may be in either direction, *i.e.,* the second practice may disturb retention of the first skill, or the initial practice may slow down the learning of the second task.

Such interference is also a function of the amount of time elapsing between the two tasks to be learned. In studies of interference with retention (retroactive inhibition), a short rest between the two kinds of practice is very effective in reducing the interference effect. Potentially interfering phases of learning in any training program

should therefore be separated by rests or by periods of completely distinct activity. Interference is also found to be greatest when the initial activity is only partially learned. We may therefore recommend that one of the two activities should be thoroughly practiced before the second activity is introduced to the trainee.

Scheduling Practice Sessions. Experiments with verbal memorizing and with the learning of motor skills have shown that the length of practice sessions is an important determiner of the effectiveness of practice. More is learned per unit of time spent in practice when a number of short periods are distributed over several days, rather than when all the practice time is massed together. The actual length of practice period which is optimal for a given task will, of course, depend upon the nature of that task. Nevertheless a rough generalization might be that the optimal length of period is usually shorter than that which would be chosen by a person who is unaware of the experimental results. Consequently there is a general tendency to err in the direction of too great length of session.

Many times the optimal spacing of practice cannot be secured because of practical demands. Distributed practice is more effective from the point of view of maximum use of *practice time,* but the practical instructor must also consider the adequate use of the remaining time of the trainees. It might be that progress per hour of practice in a given job, for example, would be greatest if the trainees spent one-half hour twice a day in working at the task. It would, however, take them several months to learn. By practicing several hours a day they are able to learn the task within a few weeks. From the point of view of total time-span spent in training, the massed practice may therefore be more effective.

Although some considerations like that given in the above example will cut across the recommendations for distribution of practice, the experimental results are not to be ignored. There are a number of ways in which the general findings are important. If, for example, a task consists of several distinct components (for example, a job which requires operating a drill press, filing, measuring, layout of work), each component can be scheduled for short practice sessions distributed over days and weeks, since the remaining time of the trainee can be used in practice upon the other components. Spacing of practice does not imply that the trainee must be idle between practice sessions.

There are even instances where it might pay to allow the trainee to work only half a day, even from the point of view of the total

time-span of the training period. For example, the writer once heard of a group of men learning radio code who practiced eight hours per day throughout the program. The men undergoing this training had already received their basic naval instruction, and were not scheduled for any other type of training at the time. Consequently it was assumed that they should spend full time in their instruction and practice upon this one skill. The optimum length of practice session for this task has not as yet been established, but what evidence we have indicates that eight hours is certainly too long for maximum effectiveness of practice time.[3] In fact it is quite possible that this period is too long from any point of view. Perhaps the total gain *per week* of the program would have been just as great if the men had spent as little as three or four hours per day in practice. It is even conceivable that the total time-span of the program could have been shorter with reduced hours of practice per day. We do not know whether or not this is true, but the example illustrates the common failure to make any examination of this question, and the assumption that improvement in skill is a uniform function of the time spent in practice.

Motivational Factors in Learning. There is a considerable body of evidence that repetition is not in itself a sufficient condition for learning. The evidence consists of experiments in which the subjects are engaged in an activity in such a way that they do not realize that the experimenter is studying their learning of that activity. Under these conditions, lack of any intention to learn the activity makes the repetition almost completely ineffective. Although there may be conditions other than the explicit intention to learn which make repetition effective, it is nevertheless clear from these experiments that the effectiveness of practice will vary more with motivational conditions than with any of the factors considered so far. Indeed, one possible explanation for the facts we have already considered, such as distribution of practice and the subdivision of material, may lie in the variations in interest which occur as a result of these factors in practice.

Although the importance of motivation cannot be overemphasized in any discussion of learning, nevertheless little can be added beyond what has already been said earlier in this book in our chapters on

[3] In a review of the literature upon code learning, Taylor mentions that two investigators have recommended periods of not more than an hour per day (upon the basis of uncontrolled observations rather than experimental study). D. W. Taylor, Learning telegraphic code, *Psychological Bulletin,* 1943, *40,* 461-487.

incentives and motives. We need only consider special phases of the problem which come up in the learning situation.

It will be noted that the experiments in which practice was ineffective did not involve lack of *any* motivation upon the part of the subjects. They were motivated to perform, but not to learn. In other words, the *direction* of motivation or intention is an important factor in the learning process. On the other hand, we do find instances in everyday experience in which we learn something without the intention to learn. We apparently learn it because we examine (perceive, comprehend, or inspect) it with interest, or we get some kind of enjoyment from the activity. In the experiments which resulted in little learning without intent to learn, the material itself had little meaning or interest to the subjects.

This brief glimpse of the motivational factors in learning brings us back once more to problems of efficiency which have been discussed before. We should like to know the relative cost of learning uninteresting material under extrinsic incentives such as pay, as compared with learning material which is intrinsically interesting. Common experience would lead us to expect that the former case requires both additional effort and time for learning. If experimental studies could show that these differences are large enough to be practically important, we should then have an answer to the question of how much the instructor should do to make the task intrinsically interesting.

The instructor in a particular training program is limited in the extent to which he can build up intrinsic interest in the work among his trainees. The ability to become interested is, after all, a function of many things which have happened to the trainees before they came to the present instructor. He can widen those limits by proper selection of men who are to enter the training program, but once the men are selected his use of the factor of interest is definitely restricted. Nevertheless he must make maximum use of this factor. Extrinsic incentives in the form of praise, reproof, or monetary bonuses can be only a partial substitute for this factor. They must be considered as such, and not given the central role to which they are sometimes assigned.

Understanding in Learning. Understanding plays a multiple part in learning. In those motor and verbal tasks where the thing to be learned is a sequence or pattern—a series of words to be reproduced verbatim, a series of movements to be reproduced in detail—understanding is not essential to the performance of the task. Even

here, however, the introduction of meaning obviously makes the task easier to learn. Other kinds of learning depend upon the comprehension of the relationships and problems to be learned.

Understanding may also provide its own incentive which leads to learning. The mere fact that something is comprehended by the individual often seems to provide it with enough interest to ensure its retention. In other words, it may be that to understand something is in itself a factor in interest. In addition, of course, understanding can relate the activity to be learned to other fields of endeavor which have already acquired interest for the learner.

To serve these functions, understanding does not have to be complete, technically detailed, or even rigorously correct in the sense in which it would be regarded by an expert in the given field. The instructor must concentrate upon achieving a method of presentation which gives an understanding at a level which is satisfactory to the trainee, and which is sufficient to serve the functions listed above. Thus the radio technician trainee does not need a complete understanding of the theory of electron emission. The conception of electron emission which he is given may be subject to serious criticism by a professional physicist, but the trainee's conceptions may be quite adequate for his purposes.

In these areas the skill of the instructor must consist in achieving a compromise between detailed and accurate technical knowledge and loose conceptions which are likely to lead to serious error. The difficulty which is met in many training programs is in finding or training instructors who know the technical basis of the job they are teaching, and yet are willing to forego complications in their instruction.

Conclusions.—The recent war brought with it a great increase in the number of individuals specializing in the direction and planning of training programs. Though many of these specialists collected valuable experiences and observations, few of the observations were made under anything like controlled conditions. Many of those who directed training were themselves untrained for the work. Those who did have some background in educational method were too often attempting to extrapolate from what they had learned of schoolroom techniques to the problems of factory training. That would be satisfactory only if experimental studies were used to check upon these skilled guesses.

This situation is apparently the reason why we have not seen very many reports of the controlled testing of training methods in industry in spite of the large amount of training which went on during the

war. Perhaps such studies have been carried out and have not yet been published for one reason or another. Yet the controlled study of training methods, with deliberate variation in the conditions of learning, with *industrial* tasks learned in the industrial setting, is the only sound way in which our techniques of training can be advanced. Once more we come back to our much repeated point that industry cannot expect decisive aid from psychology until it actively maintains research programs directed toward the solution of its particular problems, not only its immediate, everyday problems, but also the broad fundamental questions which underlie them.

INDEX OF NAMES

Adams, S., 129
American Society of Heating and Ventilating Engineers, 136
Arai, T., 67, 68
Ash, I. E., 60
Atzler, E., 38, 39, 54, 61, 110, 164

Baker, K., 131
Balchin, N., 163
Barkin, S., 222
Barmack, J. E., 42, 202
Barnes, R. M., 158, 159, 167, 213, 217, 226, 230, 231, 232
Bedford, T., 137, 264
Benedict, C. G., 41
Benedict, F. G., 35, 41
Bentley, M., 97
Berens, C., 141
Berger, C., 144
Berrien, F. K., 129
Bills, A. G., 42, 68, 69, 78, 87, 99, 100, 109
Birren, J. E., 117
Bitterman, M. E., 47, 87, 102, 141
Blankenstein, S. S., 115
Bloch, W., 228
Bolanovich, D. J., 244
Borschtschewski, A. S., 151
Brozek, J., 117
Brush, A. L., 290

Cain, P. A., 199, 200, 201, 205
Carroll, P., Jr., 215
Chambers, E. G., 273, 278, 285
Clothier, R. C., 175
Cobb, P. W., 266, 279, 280, 281, 282, 294
Cohen, L., 219, 220
Collier, R. M., 175
Cooperman, N. R., 154, 155
Courts, F. A., 104, 109
Cox, J. W., 306
Crowden, G. F., 167
Culpin, M., 190, 293

Darrow, C. W., 107, 108
Davis, R. C., 104, 105, 127
Dealey, W. L., 166

DeSilva, H. R., 270, 271, 276, 279, 282, 283
Dickson, W. J., 150, 178, 181, 192
Drinker, P., 137
Dunajewski, M. J., 111
Dunbar, H. F., 191, 290, 291, 292, 295
Dvorak, A., 166

Enzer, N., 115
Ewart, E., 244

Farmer, E., 161, 190, 270, 273, 274, 275, 278, 285, 287, 295
Fay, N. J., 112
Ferderber, M. B., 137
Féré, C., 53, 58, 60
Ferree, C. E., 140, 141, 143
Fisher, M. B., 117
Fleischer, W. L., 137
Fletcher, E. D., 272, 273, 283, 284, 294
Flinn, R. H., 111, 116, 256
Florence, P. S., 71, 72, 260
Forbes, T. W., 271, 282
Ford, G. C., 166
Fraser, J. A., 73
Freeman, G. L., 106, 107, 108
French, L., 110

Giese, W. J., 108
Giesecker, F. E., 137
Gillespie, R. D., 101
Golden, C. S., 223
Goldmark, J., 71, 260
Gordy, C. B., 233, 234, 235
Gorkin, S. D., 151
Graf, O., 80, 150, 151
Guilford, J. P., 245
Gutberlet, C., 137

Hammond, E. C., 111, 256
Harmon, F. L., 127
Hartmann, G. H., 163
Heinrich, H. W., 286
Herdan, G., 287
Hersey, R. B., 293
Hoagland, H., 112
Hoke, R. E., 166

Holcomb, R. L., 262
Hollingworth, H. L., 68
Holway, A. H., 115, 141
Hopkins, M. D., 71, 72, 260
Houghten, F. C., 137
Hunt, J. McV., 191
Hunt, T., 110
Hunter, O. B., 110
Huntington, E., 133
Hurvich, L. M., 115, 141

Ioteyko, J., 51
Irwin, O. C., 108

Jellinek, E. M., 262
Johnson, H. M., 29, 86, 88, 153, 242, 256, 262, 279, 282, 292
Jones, B. F., 111, 116, 256
Juran, J. M., 219
Jurgensen, C. E., 285

Kaplan, P. M., 111
Keys, A., 117
King, B. G., 117
Kirk, F. J., 273, 278
Kleitman, N., 152, 154, 155
Knehr, C. A., 141
Kraepelin, E., 80
Kurbatova, I. N., 56
Kustanowitsch, B. J., 151

Laird, D. A., 127, 154
Langdon, J. N., 131, 198
Lauer, A. R., 276, 279
Lawshe, C. H., 250, 252
Lee, F. S., 71, 72, 260
Lee, R. H., 114, 116
Levina, A. E., 110
Lewina, A. J., 111
Lewitina, G. A., 111
Liddell, H. S., 191
Link, H. C., 73, 74, 199, 254
Lorge, I., 228
Louden, J. K., 219
Lovekin, O. S., 102
Lowry, S. M., 214, 215, 223, 225
Luckiesh, M., 138, 139, 140, 141

McFarland, R. A., 115, 141, 262
McGeoch, J. A., 304
Mace, C. A., 171, 172, 173
Maggiora, A., 58, 60
Mathewson, S. B., 175
Maynard, H. B., 214, 215, 223, 225
Merrick, N. I., 166

Mills, C. A., 133
Moore, B. V., 163
Morgan, J. J. B., 104, 126
Moss, F. A., 110
Mosso, A., 49, 56
Mullin, F. J., 154, 155
Münsterberg, H., 203
Murschhauser, H., 35
Muscio, B., 47

Newbold, E. M., 266
Norton, H. F., 165

Offner, M., 80
Ogle, C., 133
Osborne, E. E., 264

Paterson, D. G., 144
Pincus, G., 112
Poffenberger, A. T., 28, 40, 41, 45, 46, 68, 128, 161, 185
Pollock, K. W., 233
Presgrave, R., 214, 216, 221, 223

Rand, G., 140, 141, 143
Reid, C., 56, 59
Richter, W., 150, 151
Robinson, E. S., 68
Roe, J. H., 110
Roethlisberger, F. J., 150, 178, 181, 192
Rosenberg, A. A., 137
Rosenblum, D. E., 41
Rothe, H. F., 74
Rounds, G. H., 41
Rubarth, B., 165
Ruttenberg, H. J., 223
Ryan, A. H., 82, 83, 84, 85, 86, 87
Ryan, T. A., 35, 79

Samytschkina, K. S., 111
Sappenfeld, P. R., 22
Satter, G. A., 250, 252
Schubert, H. J. P., 41
Scott, W. D., 175
Seashore, S. E., 244
Segur, A. B., 233
Shakhnovich, C. E., 110
Shapiro, S. S., 111
Sharp, L. H., 58, 59
Sheidin, Y. A., 56
Shumard, F. W., 214, 215
Sidorowa, L. M., 111
Simonson, E., 54, 56, 57, 61, 115
Simpson, R. M., 107
Skliankaia, R. M., 110

Smith, M., 81, 82, 83, 190
Snell, P. A., 114
Snow, A. J., 276
Soloway, E., 102
Spriegel, W. R., 175
Stacey, A. E., Jr., 137
Stegemerton, G. J., 214, 215, 223, 225
Stock, F. G. L., 73, 131, 198
Strauss, L., 219
Swan, T. H., 153

Tasker, C., 137
Tauber, E. S., 290
Taylor, D. W., 309
Taylor, F. W., 164
Thorndike, E. L., 67
Tiffin, J., 240, 244, 246, 265, 286
Tinker, M. A., 138, 139, 141, 142, 144
Titelbaum, S., 154, 155
Tschernomordik, O. S., 111

Uhrbrock, R. S., 218

Vernon, H. M., 73, 127, 145, 257, 258, 259, 261, 264, 278
Volmer, E., 117

Warner, C. G., 127, 264
Warner, M., 82, 83, 84, 85
Watson, G., 176, 177
Wechsler, D., 277, 278
Weigand, G. E., 153
Weiss, A. P., 276
Wenger, M. A., 108
Wenzig, K., 164
Weston, H. C., 129
Wheeler, W. L., Jr., 154
Wilson, C. H., 112
Wolfe, T. P., 290
Woodworth, R. S., 107, 304
Wyatt, S., 73, 130, 131, 185, 186, 198, 202, 204, 206, 207, 208, 214

Yaglou, C. P., 137

Zak, A. I., 110

INDEX OF SUBJECTS

Absenteeism, as an index of efficiency, 121
Absolute validity of time standards, 222-224
Accident control, psychological problems in, 253-296
Accident distribution, 268, 269
Accident proneness, 264-296; clinical use of tests for, 281-284; defined, 265; generality of trait, 284-289; personality factors in, 289-293; relation to neurosis, 293; tests for, 270 *ff.*
Accident rates, as measures of efficiency, 120; comparison with war casualties, 253; prediction by tests, 278-281
Accident types, relations of, 286-289
Accidents, alcohol as a factor in, 260-263; due to sleeping driver, 255, 256; emotional factors in, 289-293; fatigue as a factor in, 254-260; lighting as a factor in, 263; morning *vs.* afternoon hours in industry, 259; personality factors in, 289-293; reduction by educational campaigns, 254; relation of driving and other accidents, 288, 289; relation of major and minor accidents, 285, 286; temperature as a factor in, 263, 264
Achievement tests, 299
Action potentials. *See* Muscle potentials
Acuity, visual, effect of level of illumination on, 141
Alcohol, comparison of alcohol and fatigue effects, 88; effect on habitual performances, 262; relation to accidents, 260-263
Alternation of shifts, their effect on production, 149
Applied psychology, nature of, 3-7
Aptitude tests, combining scores in, 242; effort index in, 22. *See also* Driving tests
Arrangement of working place, 159
Attitudes, as a factor in efficiency, 173, 174, 186, 187. *See also* Incentives
Automobile driving, effects of, 82-87, 111, 116. *See also* Accidents

Barrow work, energy cost of, 167
Blinking, as an index of effort in visual work, 141
Blocking, as a factor in output decrement, 69
Blood count, as related to effort and fatigue, 110, 111; effect of muscular work on, 110; effect of sedentary work on, 110, 111; effect of truck operation on, 111
Blood pressure, as an index of effort, 101, 102; in sedentary work, 101, 102
Bonus payments, their effect on production, 185, 186
Boredom, 184, 196-208; effect of music on, 130, 131; energy cost of, 42; factors causing susceptibility to, 203-208; generality of trait of susceptibility, 206, 207; nature of, 197; questionnaire studies of, 202; relation of fatigue to, 197; relation to intelligence, 204; relation to physiological measures, 202; susceptibility to, 203-208

Calculation, output decrement in, 68
Capacity, changes in, related to efficiency, 89; definitions of, 30; reduced in fatigue, 43, 45-48; relation to effort, 31
Carelessness, tests for, 278
Central nervous system, in fatigue, 55
Chemical theories of fatigue, 60
Circulatory indices of effort and fatigue, effect of illumination on, 141; in sedentary work, 100-103
Clinical treatment of accident-prone drivers, 281-284
Cold, effects of. *See* Temperature
Color of illumination, effects of, 143
Comfort chart for effective temperature, 136
Comfort reports, related to temperature, 134-138
Comfort zones for temperature, 137
Conflict, as a factor in work decrement, 68-70
Contrast, visual, in working materials, 143, 144

Control, experimental, 16-19
Convergence reserve, as a measure of visual fatigue, 140, 141
Cooling. *See* Temperature
Coordination, effect of fatigue on, 60; tests in accident-proneness, 270 *ff.*
Correlation, 12-16; curvilinear, 16; linear, 13, 14; multiple, 16; significance of, 15, 16; of types of accidents, 287
Cost of work, 7, 8; definition of, 21; dissatisfaction as a factor in, 190; overhead, 40. *See also* Effort, Energy cost of work, Fatigue
Crank, effect of height and load on energy cost of work, 38
Critical levels of illumination, 139
Critical scores, 243

Day shift, compared with night shift, 148
Decrement. *See* Output decrement
Dependence of variables, 12, 16-19
Depth of fatigue, 53
Design, of tools and equipment, 159, 164-167; of typewriters, 165-167
Difference, significance of, 11
Discrimination, effect of illumination level on, 140
Disease, effect of temperature on, 133; emotional factors in, 191
Dissatisfaction with work. *See* Boredom, Job satisfaction
Distraction, effects of, 124. *See also* Noise
Distribution, of Illumination, 142; 143; of practice, 308; of rest periods, 150
Driver tests, 270-284; validity of, 270-281
Driving, permissible hours, 256, 257. *See also* Accidents, Automobile driving
Dynamic work. *See* Work, muscular

Ease of seeing, design of numbers for, 144. *See also* Illumination
Economy of work. *See* Efficiency
Educational campaigns in accident reduction, 254
Effective temperature, chart, 136; definition, 135
Efficiency, aims of study, 26-30; definition of, 7, 8, 20; effect of boredom on, 196-208; effect of incentives on, 169-187; effect of lighting on, 138-144; effect of loss of sleep on, 154; effect of methods of work on, 157-168; effect of noise on, 124-131; effect of rest periods on, 149-151; effect of temperature on, 131-138; ideal of, 28, 29; interrelations of factors in, 30, 31; misuses of term, 21; problems of, 7-9; relation to emergency, 26-28; relation to fatigue, 47; relation to muscle tension, 109; relation to pulse rate, 102; summary of the concept of, 32
Efficiency measures, absenteeism, 121; accident rates, 120; change in capacity, 89-91; health records, 121; statistical, 118-121; turnover rates, 121. *See also* Effort, Energy cost of work, Fatigue, Output
Efficiency, mechanical, computation of, 37; definition of, 20; effect of rest periods on, 151; measurement of, 35-42; of lifting by a crank, 38; of muscular work, 38-42; of pulling and pushing, 39; of walking and running, 40; relation to fatigue, 54
Efficiency, total, 29, 30
Effort, concept of, 95-98; definition of, 22; effect of practice on, 297; rating of, 215, 216; relation to incentives, 169; relation to time standards, 210
Effort measures, blood count, 110, 111; circulatory, 100-103; in visual work, 138-142; ketosteroid output, 112, 113; muscle tension, 103-109; pulse product, 102; pulse rate, 100; physiological, 95-123
Effort rating, in time standards, 215, 216
Electrical resistance of the skin, as an index of tension, 106-109
Emergency conditions, efficiency in, 26-28
Emotional factors, in accidents, 290-293; in disease, 191; related to fatigue, 188-208
Emotional stability, in accidents, 276, 277
Energy cost of work, 21, 35-42; effect of height and load in lifting by a crank, 38; effect of rest, 151; effect of shovel design on, 164; of barrow work, 167; of sedentary work, 41-42; of work during noise, 127-128
Ergograph, 49
Ergographic experiments, results of, 51-60
Errors, related to efficiency, 120
Exhaustion, 52; sudden, following slow work, 60
Exit interview, 174

Factorial analysis, of accident types, 287
Factory studies, of effect of noise, 129, 130; of incentives, 173-176; of lighting, 139; of rest periods, 149-151. *See also* Efficiency, Output, Output curves
Factory work, incentives in, 169-187; output decrement in, 70-75; repetitious, 196-208
Fatigue, and mechanical efficiency, 54; as a cause of accidents, 254-260; cyclical character of recovery from, 58; defined as reduced efficiency, 47; definitions of, 42, 43; depth of, 53; effect of, on the blood count, 110, 111; emotional factors in, 188-208; method of computing decrement, 67; output decrement in factory work, 70; physiological basis of, 44; relation to alcohol effects, 88; relation to boredom, 197; transfer of, 78, 87, 88
Fatigue allowances, 213, 225-227; table of, 226
Fatigue decrement. *See* Output decrement
Fatigue in muscular work, 48-62; chemical basis of, 60, 61; effect of load on rate of development of, 60; effect of rate of work on, 60; effect of, on coordination, 60; lactic acid in, 61; loci of, 55-57; spread of activity in, 60
Fatigue in sedentary work, 63-95; decrement resulting from, 65-75; methods of study of, 75-78
Fatigue measures, flicker-fusion frequency, 113-118; ketosteroid output, 112-113; physiological tests, 95-123; reduced capacity in, 43, 45-48; reduced efficiency in, 47; work decrement, 48, 65-75
Fatigue, nervous, 188-208
Fatigue tests, 76-88; effects of automobile driving on, 82-87; effects of sleep-loss on, 81-83, 153; in finding permissible hours of driving, 256, 257
Flicker-fusion frequency, 113-118; effect of exposure to flicker on, 114, 115; effect of heavy muscular work on, 117; effect of low oxygen pressure on, 117; effect of sleep-loss on, 115; effect of truck driving on, 116
Focusing, speed of, as affected by illumination, 140
Food, effect of, on sleep, 154-156
Force, amount required as a criterion of typewriter design, 165; dependent on handle design, 165; mechanical, in muscular work, 33
Fuel consumption. *See* Energy cost of work

Generality of traits, accident-proneness, 284-289; susceptibility to boredom, 206-207
Glare, effect of, 138, 142-143
Goal, effect of, on output, 171-173. *See also* Time standards
Governing factors, 124-156 relation to incentives, 124
Graphic rating method, 238-246

Halo effect, in industrial merit rating, 244
Handles, design of, for screwdriver, 165
Health, effect of temperature on, 133; emotional factors in, 191; records as a measure of efficiency of, 121
Heart rate. *See* Circulatory indices, Pulse rate
Heat, effects of. *See* Temperature
Homogeneity, as a factor in work decrement, 68-70; relation of, to motion study, 161
Hours of work, 145-149; effect on accidents, 257-259; permissible, in truck operation, 256, 257
Humidity, effect of, 132

Illumination, color of, 143; critical levels of, 139; effect of level of, on acuity, 143; effect of level of, on rate of blinking, 141; effect of level of, on speed of discrimination and focusing, 140; glare, 142-143; indirect, 143; levels of, 138-142; methods of determining optimal levels of, 138-142; preferences for, 141-142; relation of, to accidents, 263
Impairment of efficiency, relation to fatigue tests, 86
Impairment of output. *See* Output decrement
Incentives, 8, 169-187; effect of time standards on, 209, 224; in learning, 299, 309, 310; methods of study of, 170-176; pay, 8, 9, 175, 183-186, 214; relation of, to efficiency, 169; relation of, to governors, 124, 169; relation of, to merit rating and job evaluation, 237; results of questionnaire studies on, 176-178; social factors of, 186;

Incentives (*cont'd*)
standards and goals for, 171-173, 182; Western Electric Company studies of, 150, 178-182
Incentives, industrial, field studies of, 173-176; naturalistic studies of, 178
Individual differences, in accident liability, 264-296; in studies of efficiency, 39; relation of, to motion study, 162, 163; relation of, to synthetic time standards, 233-235; relation of, to time standards, 213, 214; susceptibility to monotony, 196, 203-206. *See also* Accident proneness
Input, definition of, 21. *See* Cost of work, Effort, Energy cost of work
Intelligence, relation of, to accidents, 270, 278; relation of, to boredom, 203, 204; relation of, to turnover rate, 175
Intensity of light. *See* Illumination, levels of; Lighting
Intent to learn, 309
Interest, 183-186; as a factor in learning, 310
Interference in learning, 305-308
Intersection accidents, and vision, 276, 294
Interview studies of incentives, 174, 179

Job analysis. *See* Job evaluation, Methods of work, Motion study
Job evaluation, 209, 237, 238, 250-252; factors in, 251; halo effect in, 252
Job hazards, differences in, 265
Job preferences, 177
Job satisfaction, 177, 178, 184

Ketosteroids, as an index of fatigue, 112, 113

Lactic acid, role of, in fatigue, 61
Learning, adjustive, 302 *ff.*; by wholes, 304; effect of sequence on, 303-308; effect of, on muscle tension, 109; efficiency of, 297-312; incentives and motives in, 309-310; interest as a factor in, 310; interference, 305-308; kinds of, 301-303; memorizing, 302 *ff.*; motor, 302 *ff.*; relation of, to comprehension, 302 *ff.*; sequence of material, 303, 304; understanding as a factor in, 310, 311
Legibility of type faces, 144
Length of rest periods, effect of, 58, 150
Length of working day, effect of, on accidents, 259; effect of, on productivity, 146
Leucocytosis. *See* Blood count
Leveling factors, table of, 215
Leveling method. *See* Time standards with rating
Levels of effort, criteria of, 223
Lighting, effects of, 138-144. *See also* Illumination
Load, effect of, on efficiency, 38, 39; effect of, on fatigue, 60; optimal, for shoveling, 164; relative, on two hands in typing, 166
Locomotion, mechanical efficiency and rate of, 40

Machine work, output curves in, 70-75
Machines, design of, 164-167
Major and minor accidents, relations of, 285, 286
Mechanical efficiency. *See* Efficiency, mechanical, Energy cost of work
Mechanical recording, of muscle tension, 106
Memorizing, factors in, 302 *ff.*
Mental work. *See* Work, sedentary
Merit rating, 237-250; correcting for differences of standards, 245; halo effect in, 244 *ff.*; reliability of, 240 *ff.*; suggested procedure for, 246-249; validity of, 240 *ff.*
Metabolic cost of work. *See* Energy cost of work
Methodology, in applied psychology, 9-19; of setting time standards, 210-217
Methods engineering. *See* Methods of work, Motion study
Methods of work, 157-168; arrangement of tools and material, 159; design of tools, 159, 164-167; principles of, 158, 159; relation to individual differences, 162, 163; use of the body in, 158
Micromotion chart, 158
Monotony. *See* Boredom
Morale, 169
Motion economy. *See* Methods of work
Motion study, 157-168; criticisms of, 160-164; relation to efficiency, 160. *See also* Methods of work
Motivation, problems of, 8. *See also* Incentives
Motor learning, factors in, 302 *ff.*
Motor skill, 64; concept of, in time standards, 211, 212. *See also* Learning
Muscle potentials, 104-106; in rest, 58, 59
Muscle tension, as an index of effort, 103-109; effect of learning on, 109;

effect of lighting on, 141; effect of noise on, 104, 126-128; effect of, on output, 109; relation of, to efficiency, 109; relation of, to electrical skin resistance, 106-109; role of, in sedentary work, 99, 100, 103-109
Muscular work. *See* Work, muscular
Music, effects of, 130, 131; set as a factor in determining effect of, 131

Nervous fatigue, 188-208
Neuroses, experimental, 191; occupational, 190; relation of, to accident-proneness, 293
Night shifts, effect of, on productivity, 148
Noise, effect of, on efficiency and productivity, 124-131; effect of, on muscle tensions, 104, 126-128; effect of, on the energy cost of work, 127, 128; field studies of its effects in industry, 129, 130

Ocular work, factors in, 138-144
Output, effects of boredom on, 197-202; effect of emotional factors on, 188, 189; effect of illumination on, 139; effect of loss of sleep on, 153; effect of muscle tension on, 109; effect of noise on, 125, 129; effect of rest periods on, 149; effect of temperature on, 133; measurement of, in foot-pounds, 36; relation of, to capacity, 47; relation of, to tiredness, 46; restriction of, 175, 176; under the piece-rate system, 185, 186, 214; units of measurement of, 37, 38
Output curves, in boring work, 197-201; in factory work, 70-75; in inspection work, 73, 74, 199
Output decrement, as a fatigue measure, 45-48; effect of conflict in task on, 68-70; effect of homogeneity of task on, 68-70; factors in, 65-70; in factory work, 70-75; in a packing operation, 74; in sedentary work, 48, 65-75; method of computing, 67; related to fatigue allowances, 225
Output rate, relation to accidents, 258-260
Output trends, relation to efficiency, 119
Overlapping motions, in synthetic time standards, 233
Oxygen consumption, as a measure of the energy cost of work, 35
Oxygen debt, 36
Oxygen deprivation, effects of, on flicker-fusion frequency, 117

Part *vs.* whole learning, 304
Pay incentives, relation of, to motion study, 161. *See also* Incentives
Perceived fatigue. *See* Tiredness
Performance rate. *See* Output
Personal factors. *See* Accident-proneness; Boredom, susceptibility to; Individual differences
Personality, and accidents, 289-293; and boredom, 203-208
Physiological effects, of boredom, 202; of sleep, 152; of work, 95-123
Piece-rate methods, effect of, on production, 185, 186, 214
Practice, effect of, on effort, 297; effective, 297-312; scheduling of, 299, 308, 309; sequence of, 303-308; size of unit of, 304, 305. *See also* Learning
Practice effect, control of, in measuring fatigue decrement, 67
Prediction, 12-16
Problem-solving, 64
Process chart, 158
Production rate. *See* Output
Propaganda, effect of, on accident rates, 254; methods of evaluating, 254
Psychiatric study of accident-proneness, 289-293
Psychology, applied, the nature of, 3-7
Pulling *vs.* pushing, relative energy cost of, 39
Pulse-product, as an index of effort, 102
Pulse rate, as an index of effort, 100-102; in visual tasks, 102

Questionnaire, studies of boredom, 202; survey of incentives, 173, 174, 176-178

r (symbol for linear correlation). *See* Correlation
Radiation, as a factor in temperature control, 132
Rate of work, problems of, 26-28; effect of, on fatigue and output, 53, 60
Rate setting. *See* Job evaluation, Time standards
Rating methods, in time study, 213-236. *See also* Merit rating, Job evaluation
Reaction time, in accident-proneness, 270 *ff.*
Recollection, 301

Recovery, cyclical character of, 58; muscle tension during, 58, 59. *See also* Rest
Reliability, of merit ratings, 240 *ff.*; of time standards, 217-221
Repetitive work, effects of. *See* Boredom
Research methods, general, 9-19
Resistance of skin. *See* Electrical resistance of skin
Rest, following muscular work, 57-59; muscle potentials during, 58, 59
Rest periods, 145, 149-151; effect on pulse-product, 103; in sedentary work, 150, 151; optimal length, 150; Western Electric Company studies of, 150
Restriction of output, 175, 176
Retroactive inhibition, 307, 308
Running, mechanical efficiency of, 40

Safety propaganda, problem of effectiveness of, 254
Sampling error, in correlation, 15, 16
Satisfaction with work. *See* Job satisfaction
Scatter diagram, 14
Scheduling of practice, 299, 308, 309
Screwdrivers, design of, 165
Sedentary-muscular work, 64
Sedentary work. *See* Work, sedentary
Selection, vocational, 8, 9
Sensory adjustment, 64
Sequence, effect of, on learning, 299, 305-308
Shifts, alternating, 149; day and night, 148
Shovels, design of, 164; effect of design of, on energy cost of work, 164, 165; optimal load of, 164, 165
Significance, of correlation, 15, 16; practical, 10-12; statistical, 10-12, 15, 16
Skill, development of, 297-312; relation of, to time standards, 211. *See also* Learning
Skin resistance. *See* Electrical resistance of skin
Sleep, 145, 151-156; criteria of, 152; distribution of, 152; effect of food on, 154-156; effect of loss of, on energy cost and efficiency, 154; effect of sleep loss on test performance, 81-83, 153; movement during, 154; physiological changes during, 152
Sleeping drivers, as a cause of accidents, 255, 256
Social factors in work, 186
Spacing, of muscular contractions, effect of, 60
Speed of work. *See* Output rate
Spread of muscular activity in fatigue, 60
Standards, effect of, on output, 171-173, 183; methods of determining. *See* Time standards
Static work. *See* Work, muscular
Statistical measures, of efficiency, 118-121
Statistical methods, 10-16; of setting time standards, 228
Supervisory relations, 186
Susceptibility to boredom. *See* Boredom
Synthetic time standards, 229-235; relation to individual differences, 233-235

Telegrapher's cramp, 190
Telegraphic code, learning of, 309
Temperature, comfort zones, 137; effect of, 131-138; effect of, on health, 133; methods of determining optimum, 132-138; physiological effects of, 132; relation to accidents, 263, 264
Tests, achievement, 299. *See also* Accident proneness, Fatigue tests
Therblig, definition of, 158
Time standards, 209-236; leveling method for, 213-236; methods of setting, 210-217; rating methods for, 213-236; relation of, to average time, 214, 228-229; relation of, to individual differences, 213, 214; reliability of, 217-221; statistical methods of setting, 228; synthetic methods for, 229-235; traditional, 212, 213; use of several rates of, for improving reliability, 219; validity of, 221-227
Time study, as a method of determining the design of equipment, 164. *See also* Time standards
Tiredness, feelings of, 43-45; in relation to output, 46
Tonus. *See* Muscle potentials, Muscle tension
Tools, design of, 159, 164-167
Training, effective, 297-312; of merit raters, 244
Transfer, of fatigue, *See* Fatigue; of training, 305-308
Truck driving. *See* Accidents, Automobile driving
Turnover rate, analysis of, 174; as in-

dicating efficiency, 121; relation of, to intelligence, 175
Types of accident, interrelations of, 286-289
Typewriters, design of, 165, 166

Understanding, as a factor in learning, 310, 311
Uniformity of task, its effect on decrement, 68-70. *See also* Boredom

Validity, of driver tests, 270-281; of merit ratings, 240 *ff.*; of time standards, 221-227
Ventilation, effects of, 131-138
Visibility, of numbers, 144
Visual tests, in accident-proneness, 270 *ff.*; relation to intersection accidents, 276, 294
Visual work, factors in, 138-144
Vocational selection, 8, 9

Wages. *See* Incentives
Walking, mechanical efficiency of, 40
Warming-up, effect on work-curves, 68
Weighting, of traits in rating, 243 *ff.*
Western Electric Company researches, 150, 178-182, 192
Wheelbarrow, design of, 167; work with, energy cost of, 167
White cells. *See* Blood count
Whole learning, 304
Work, classification of, 33-35; dromal tasks, 57; dynamic, fatigue in, 55, 56; static, fatigue in, 56, 57
Work, muscular, definition of, 33; effect on blood count, 110; effect on fatigue, 48-62; effect on flicker-fusion frequency, 117; mechanical efficiency of, 38-42; static *vs.* dynamic, effects of, 55-57; types of, 35
Work, sedentary, circulatory indices of effort in, 100-103; definition of, 33; effect of noise on, 124-131; effect of rest periods on, 150; effect of temperature on, 133; effect on blood count, 110, 111; effect on flicker-fusion frequency, 113-118; effect on pulse-rate and pulse-product, 100-102; energy cost of, 41, 42; fatigue in, 63-95; kinds of, 63-65; muscle tension during, 103-109; output decrement in, 48; physiological correlates of, 95-123; rest periods in, 150, 151; role of muscular activity in, 99, 100
Working week, optimal length of, 146

Date Due

APR 8 '55
APR 22 '55
OCT 29 '56
MAY 17 '57
MAR 23 '59
MAR 4 '60
MAR 8 '67
[redacted]
'68
AUG 20 '72 AUG 15 '72
NOV 25 1980 NOV 23 [redacted]
FEB 10 1993 DEC 17 1992

UNIVERSITY OF FLORIDA
3 1262 04810 8327
625.5
R989w
ENGINEERING
SCIENCES
LIBRARY